Pre-Publication Copy

ICCVAM TEST METHOD EVALUATION REPORT

IN VITRO CYTOTOXICITY TEST METHODS FOR ESTIMATING STARTING DOSES FOR ACUTE ORAL SYSTEMIC TOXICITY TESTING

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences
National Institutes of Health
U.S. Public Health Service
Department of Health and Human Services

[This Page Intentionally Left Blank]

ICCVAM TEST METHOD EVALUATION REPORT

IN VITRO CYTOTOXICITY TEST METHODS FOR ESTIMATING STARTING DOSES FOR ACUTE ORAL SYSTEMIC TOXICITY TESTING

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Toxicology Program
P.O. Box 12233
Research Triangle Park, NC 27709

October 2006
NIH Publication No.: ____

National Institute of Environmental Health Sciences
National Institutes of Health
U.S. Public Health Service
Department of Health and Human Services

This Report may be cited as:

ICCVAM. 2006. ICCVAM Test Method Evaluation Report - *In Vitro* Cytotoxicity Test Methods For Estimating Starting Doses For Acute Oral Systemic Toxicity Testing. NIH Publication No. XX-XXXX (pending). National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/

Printed: _____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

3T3	BALB/c mouse fibroblasts, clone A31
ADME	Absorption, distribution, metabolism, excretion
ANOVA	Analysis of Variance
ATC	Acute Toxicity Class
ATWG	Acute Toxicity Working Group
BRD	Background review document
CASRN	Chemical Abstracts Service Registry Number
CPSC	U.S. Consumer Product Safety Commission
CS	Calf serum
CV	Coefficient of Variation
^{o}C	Degrees Celsius
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's Phosphate Buffered Saline
ECBC	U.S. Army Edgewood Chemical and Biological Center
ECVAM	European Center for the Validation of Alternative Methods
EDIT	Evaluation-guided Development of New *In Vitro* Tests
EPA	U.S. Environmental Protection Agency
ETOH	Ethanol
FAL	FRAME Alternatives Laboratory
FDA	U.S. Food and Drug Administration
FL	Fluorescein leakage
FR	Federal Register
FRAME	Fund for Replacement of Animals in Medical Experiments
GHS	Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005).

HPV	High Production Volume
IC_{50}	Test substance concentration producing 50% inhibition of the endpoint measured
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
IIVS	Institute for In-Vitro Sciences
ILS	Integrated Laboratory Systems
LD_{50}	Lethal dose that produces lethality in 50% of test animals
LDH	Lactate dehydrogenase
MEIC	Multicentre Evaluation of *In Vitro* Cytotoxicity
MTT	[3-(4,5,dimethylthiazol-2yl)2,5-diphenyl tetrazolium bromide]
NCS	Newborn calf serum
NHK	Normal human epidermal keratinocytes
NICEATM	National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NR	Neutral red
NRR	Neutral red release
NRU	Neutral red uptake
NTP	U.S. National Toxicology Program
OD	Optical density
OECD	Organisation for Economic Cooperation and Development
PC	Positive control
QSAR	Quantitative Structure Activity Relationship
RC	Registry of Cytotoxicity
RTECS	Registry of Toxic Effects for Chemical Substances
SACATM	Scientific Advisory Committee on Alternative Toxicological Methods
SLS	Sodium lauryl sulfate

SMT	Study Management Team
TESS	Toxic Exposure Surveillance System
UDP	Up and Down Procedure
UN	United Nations
VC	Vehicle control
XTT	[Sodium 3,3,-[(Phenylamino)carbonyl]-3,4-Tetrazolium-Bis(4-methoxy-6-nitro)benzenesulfonic acid hydrate]
ZEBET	German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments

INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION OF ALTERNATIVE METHODS - DESIGNATED AGENCY REPRESENTATIVES

Agency for Toxic Substances and Disease Registry
• Moiz Mumtaz, Ph.D.

Consumer Product Safety Commission
• Marilyn L. Wind, Ph.D. (Vice-Chair)
* Patricia Bittner, M.S.
* Kailash C. Gupta, D.V.M., Ph.D. (retired)
* Kristina Hatlelid, Ph.D.

Department of Agriculture
• Jodie Kulpa-Eddy, D.V.M.
◊ Elizabeth Goldentyer, D.V.M.

Department of Defense
• Robert E. Foster, Ph.D.
◊ Patty Decot
* Harry Salem, Ph.D.

Department of Energy
• Michael Kuperberg, Ph.D.
◊ Marvin Stodolsky, Ph.D.

Department of the Interior
• Barnett A. Rattner, Ph.D.
◊ Sarah Gerould, Ph.D.

Department of Transportation
• George Cushmac, Ph.D.
◊ Steve Hwang, Ph.D.

Environmental Protection Agency
Office of Science Coordination and Policy
• Karen Hamernik, Ph.D.
Office of Research and Development
◊ Julian Preston, Ph.D.
* Suzanne McMaster, Ph.D.
OECD Test Guidelines Program
* Jerry Smrchek, Ph.D.
Office of Pesticides Programs
* Amy Rispin, Ph.D.
* Deborah McCall

Food and Drug Administration
• Leonard M. Schechtman, Ph.D. (Chair)
Office of Science and Health Coordination
◊ Susanne Fitzpatrick Ph.D., D.A.B.T.
Center for Drug Evaluation and Research
* Abigail C. Jacobs, Ph.D.
Center for Devices and Radiological Health
* Raju Kammula, D.V.M., Ph.D., D.A.B.T.
* Melvin E. Stratmeyer, Ph.D.
Center for Biologics Evaluation and Research
* Richard McFarland, Ph.D., M.D.
* Ying Huang, Ph.D.
Center for Food Safety and Nutrition
* David G. Hattan, Ph.D.
* Robert L. Bronaugh, Ph.D.
Center for Veterinary Medicine
* Devaraya Jagannath, Ph.D.
* M. Cecilia Aguila, D.V.M.
National Center for Toxicological Research
* William T. Allaben, Ph.D.
Office of Regulatory Affairs
* Lawrence A. D'Hoostelaere, Ph.D.

National Cancer Institute
• Alan Poland, M.D.
◊ T. Kevin Howcroft, Ph.D.

National Institute of Environmental Health Sciences
• William S. Stokes, D.V.M., D.A.C.L.A.M.
◊ John R. Bucher, Ph.D., D.A.B.T.
* Rajendra S. Chhabra, Ph.D., D.A.B.T
* Jerrold J. Heindel, Ph.D.

National Institute for Occupational Safety and Health
• Paul Nicolaysen, V.M.D.
◊ K. Murali Rao, MD, Ph.D.

National Institutes of Health
• Margaret D. Snyder, Ph.D.

National Library of Medicine
• Vera Hudson, M.S.
◊ Jeanne Goshorn, M.S.

Occupational Safety and Health Administration
• Surender Ahir, Ph.D.

• Principal Agency Representative
◊ Alternate Principal Agency Representative
* Other Designated Agency Representatives

ACKNOWLEDGMENTS

The following individuals are acknowledged for their contributions and oversight of the in vitro acute toxicity test method evaluation process.

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Acute Toxicity Working Group (ATWG)

Consumer Product Safety Commission
Kailash Gupta, D.V.M, Ph.D. (retired)
Cassandra Prioleau, Ph.D.
Marilyn Wind, Ph.D. (ATWG Chair, ICCVAM Vice Chair)

Department of Energy
Po-Yung Lu, Ph.D.

Environmental Protection Agency
Karen Hamernik, Ph.D.
Masih Hashim, Ph.D.
Marianne Lewis, Ph.D.
Elizabeth Margosches, Ph.D.
Deborah McCall
John Redden, Ph.D.
Amy Rispin, Ph.D.

Food and Drug Administration
Leonard Schechtman, Ph.D. (ICCVAM Chair)
Kenneth Hastings, Ph.D.
Abigail Jacobs, Ph.D.
Suzanne Morris, Ph.D.
David Morse, Ph.D.
Thomas Umbreit, Ph.D.

National Institute for Occupational Safety & Health
Steven Reynolds, Ph.D.

National Institute of Environmental Health Sciences
Rajendra Chhabra, Ph.D.
William Stokes, D.V.M., D.A.C.L.A.M.
Raymond Tice, Ph.D.

European Centre for the Validation of Alternative Methods (ECVAM) Liaisons
Silvia Casati, Ph.D.
Pilar Prieto, Ph.D.

In Vitro Acute Toxicity Peer Review Panel

David H. Blakey, D.Phil., Health Canada, Ottawa, Ontario, Canada
June Bradlaw, Ph.D., International Foundation for Ethical Research (IFER), Rockville, Maryland
Robert Copeland, Ph.D., Howard University College of Medicine, Washington, DC
Gianni Dal Negro, D.V.M., Ph.D., GlaxoSmithKline Medicine Research Centre, Verona, Italy
Marion Ehrich, Ph.D., RPh., D.A.B.T. Virginia-Maryland Regional College of Veterinary Medicine, Blacksburg, Virginia
Eugene Elmore, Ph.D., University of California, Irvine, Irvine, California
Benjamin Gerson, M.D., Thomas Jefferson University School of Medicine, Philadelphia, Pennsylvania
Michael Greene, Ph.D., U.S. Consumer Product Safety Commission, Bethesda, MD
Janice Kuhn, Ph.D., D.A.B.T., Stillmeadow Inc., Sugar Land, Texas
Daniel Marsman, D.V.M., Ph.D., D.A.B.T., Procter & Gamble Company, Cincinnati, Ohio
Andrew Rowan, Ph.D., Humane Society of the United States, Washington, DC
Hasso Seibert, Ph.D., University Medical School Schleswig-Holstein, Kiel, Germany
Nigel Stallard, Ph.D., The University of Warwick, Coventry, United Kingdom
Katherine Stitzel, D.V.M., (Panel Chair), Consultant, West Chester, Ohio
Shinobu Wakuri, MSc., Hatano Research Institute, Japan
Daniel Wilson, Ph.D., D.A.B.T., The Dow Chemical Company, Midland, Michigan

National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods

David Allen, Ph.D.
ILS, Inc.

Bradley Blackard, M.S.P.H.
ILS, Inc.

Sue Brenzel
ILS, Inc.

Thomas Burns, M.S.
ILS, Inc.

Patricia Ceger, M.S.
ILS, Inc.

Jeff Charles, Ph.D., M.B.A., D.A.B.T.
ILS, Inc.

Neepa Choksi, Ph.D.
ILS, Inc.

Frank Deal, M.S.
ILS, Inc.

Linda Litchfield
ILS, Inc.

Deborah McCarley
NIEHS

Michael Paris
ILS, Inc.

William Stokes, D.V.M., D.A.C.L.A.M. (Director)
NIEHS

Judy Strickland, Ph.D., D.A.B.T.
ILS, Inc.

Raymond Tice, Ph.D. (Deputy Director)
NIEHS

James Truax, M.A.
ILS, Inc.

Doug Winters, M.S.
ILS, Inc.

Participants in the *In Vitro* Cytotoxicity Validation Study

BioReliance Corp.
(Chemical Distribution)
Dr. Martin Wenk – Principal Investigator

Institute for In Vitro Sciences (IIVS)
(Lead Laboratory – Protocols)
Mr. Hans Raabe – Study Director
Mr. Greg Mun – Laboratory Manager
Ms Angela Sizemore – Research Technician
Mr. Gregory O. Moyer – Research Technician
Dr. John Harbell – Scientific Director

U.S. Army Edgewood Chemical and Biological Center (ECBC)
(Testing Laboratory)
Dr. Cheng Cao – Study Director
Ms Janna Madren-Whalley – Research Technician
Dr. Chundakkadu Krishna – Research Technician
Dr. James J. Valdes – Scientific Advisor

FRAME Alternatives Laboratory (FAL)
University of Nottingham, UK
(Lead Laboratory – Software)
Dr. Richard Clothier – Study Director
Ms Nicola Bourne – Research Technician
Ms Monika Owen – Research Technician
Ms Rachel Budworth – Research Technician

Constella Group (Statistical Analyses)
Dr. Patrick Crockett
Dr. Eric Harvey
Mr. Robert Lee
Ms Janine Wilcox
Mr. Wendell Jones

Statistical Consultant
Dr. Joseph Haseman

NIEHS
Dr. Grace Kissling – Contract Project Officer
Ms Molly Vallant - Contract Project Officer

Study Management Team (SMT)
NICEATM
Dr. William Stokes (NIEHS) – Director
Dr. Raymond Tice (NIEHS) – Deputy Director – Advisor
Dr. Judy Strickland (ILS, Inc.) – Project Coordinator
Mr. Michael Paris (ILS, Inc.) – Asst. Project Coordinator
Dr. Jeffrey Charles (ILS, Inc.) – Advisor

ECVAM
Dr. Thomas Hartung – Head of Unit
Dr. Silvia Casati – Task Leader
Dr. Michael Balls – Head of Unit (until June 2002)

PREFACE

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) is charged by the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 [2000]; available at http://iccvam.niehs.nih.gov/about/PL106545.pdf) with evaluating the scientific validity of new, revised, and alternative toxicological test methods applicable to U.S. Federal agency safety testing requirements. ICCVAM is required to also provide recommendations to U.S. Federal agencies regarding the usefulness and limitations of test methods following their scientific evaluation.

ICCVAM initiated a review of the validation status of *in vitro* methods for estimating acute oral toxicity in 1999 in response to a request from the U.S. Environmental Protection Agency (EPA) Office of Pesticides, Prevention, and Toxic Substances. The request was based on recently published studies that showed a correlation between *in vitro* and *in vivo* acute toxicity. An International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity organized by ICCVAM and the National Toxicology Program (NTP) Center for the Evaluation of Alternative Toxicological Methods (NICEATM) was held in October 2000. Workshop participants concluded that the proposed *in vitro* methods had not yet undergone adequate studies to determine if they could meet regulatory requirements for acute toxicity testing. However, an *in vitro* approach previously proposed by the German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET) was recommended by workshop participants as a high priority for validation studies to ICCVAM (ICCVAM 2001a). The proposal was to use *in vitro* cytotoxicity data to estimate starting doses for in vivo acute toxicity studies. Since a correlation between IC_{50} and LD_{50} values had been determined based on retrospective literature reviews, such a strategy might reduce the use of animals for acute oral toxicity tests by identifying a starting dose closer to the LD_{50}. A *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (ICCVAM 2001b) was subsequently prepared by some of the workshop participants with the assistance of ICCVAM and NICEATM to provide interim *in vitro* cytotoxicity protocols and instructions for implementing the approach,

ICCVAM agreed with the workshop participants that *in vitro* basal cytotoxicity test methods should have a high priority for validation studies. The NICEATM collaborated with the European Centre for the Validation of Alternative Methods (ECVAM), to further characterize the usefulness and limitations of *in vitro* cytotoxicity assays as predictors of starting doses for rodent acute oral toxicity test methods. NICEATM and ECVAM designed an international, multi-laboratory validation study to evaluate the performance of two standardized *in vitro* neutral red uptake (NRU) test methods, using the ZEBET approach based on the Registry of Cytotoxicity (RC) regression model. One test method used BALB/c 3T3 mouse fibroblasts (3T3) while the other used normal human epidermal keratinocytes (NHK).

The validation study, which used 72 reference substances in a phased approach, was initiated in August 2002 and completed in January 2005. Upon completion, NICEATM, in coordination with the ICCVAM Acute Toxicity Working Group (ATWG) and ICCVAM, prepared a comprehensive draft background review document (BRD) reviewing the procedures and results generated from the validation study. ICCVAM subsequently convened an international independent Peer Review Panel (hereafter, Panel) meeting on May 23, 2006, to review the BRD, to evaluate the extent to which the BRD addressed established validation and acceptance criteria, and to provide comments on the draft ICCVAM recommendations on test method uses, future studies, draft test method protocols, and draft performance standards. During the Panel meeting, public attendees were allowed an opportunity to provide comment to the Panel. Public comments were also solicited through the publication of a *Federal Register (FR)* notice (Vol. 71, No. 132, pp. 39122-39123) announcing the availability of the Panel report. The draft BRD, the Panel report, and all public comments were then made available to the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM)[1], for their consideration. The SACATM agreed with the consensus conclusions of the Panel (SACATM 2006).

[1] The SACATM advises the ICCVAM, NICEATM, and the Director of the NIEHS on priorities and activities related to the development, validation, scientific review, regulatory acceptance, implementation, and national and international harmonization of new, revised, and alternative toxicological test methods.

ICCVAM and the ATWG considered the Panel report, public comments, and the comments of SACATM in preparing the final test method recommendations provided in this report. Briefly, ICCVAM recommends that, while the two standardized *in vitro* test methods (3T3 and NHK NRU test methods) are not sufficiently accurate to predict acute oral toxicity for the purposes of hazard classification, they can be used in a weight-of-evidence approach to determine the starting dose for the current acute oral *in vivo* toxicity protocols. Such use should be considered and applied where appropriate before testing is conducted using animals to reduce the number of animals needed for acute oral toxicity testing and, in some situations, to also reduce the numbers of animals that die or need to be humanely killed.

In accordance with the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 [2000]) (available at http://iccvam.niehs.nih.gov/about/PL106545.pdf), this report will be made available to the public and provided to U.S. Federal agencies for consideration. Agencies with applicable testing regulations and/or guidelines are required by law to respond to ICCVAM within 180 days after receiving the recommendations. These responses will be made available to the public on the ICCVAM website (http://iccvam.niehs.nih.gov) as they are received.

Acknowledgments

The efforts of many individuals who contributed to the preparation, review, and revision of this report are gratefully acknowledged. We especially recognize the members of the Panel for their thoughtful evaluations and generous contributions of time and effort. Special thanks are extended to Dr. Katherine Stitzel for serving as the Panel Chair. The efforts of the ATWG were invaluable for assuring a meaningful and comprehensive review. We thank the chair of the ATWG, Dr. Marilyn Wind (Consumer Products Safety Commission [CPSC]) for her effective leadership of this group. The efforts of the NICEATM staff in preparing the BRD, organizing the Panel meeting, and preparing this final report are greatly appreciated. Finally, we want to acknowledge the invaluable contributions of the laboratory staff and study directors for the validation study, the international validation study management team, and the project coordinators for the independent validation study, Dr. Judy Strickland and Michael Paris. This was the first joint validation study by NICEATM and ECVAM, and we

want to thank all of the team for ensuring excellent international coordination and communication. The experiences gained from this international cooperation are already facilitating a recently initiated second collaborative validation study with ECVAM, which also includes the new Japanese Center for the Validation of Alternative Methods (JaCVAM). International collaboration by these three centers of validation excellence will ensure high quality validation studies and take advantage of broad international expertise and experience with scientific validation.

William S. Stokes, D.V.M. Diplomate A.C.L.A.M.
CAPT, U.S. Public Health Service
Director, NICEATM
Executive Director, ICCVAM

Leonard Schechtman, Ph.D.
Deputy Director for Washington Operations
National Center for Toxicological Research

U.S. Food and Drug Administration
Chairman, ICCVAM

EXECUTIVE SUMMARY

This Test Method Evaluation Report (TMER) describes an evaluation by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) of the use of *in vitro* basal cytotoxicity test methods for estimating starting doses for acute oral toxicity tests. This evaluation provides validation information that should be helpful to various stakeholders (e.g., applicable U.S. Federal regulatory agencies, the international regulatory community, the pesticide and other commercial chemical industries) in determining when these test methods might be useful for specific testing situations. Appropriate use of these *in vitro* test methods is expected to further reduce and refine animal use for acute oral toxicity testing.

An international, multi-laboratory validation study for the use of two *in vitro* neutral red uptake (NRU) test methods was organized by the National Toxicology Program (NTP) Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the European Centre for the Validation of Alternative Methods (ECVAM) to evaluate their usefulness and limitations. In the validation study, three laboratories tested 72 reference substances for cytotoxicity in BALB/c 3T3 mouse fibroblasts (3T3) and normal human epidermal keratinocytes (NHK). The resulting data was used to estimate starting doses for rodent acute oral toxicity testing, based on linear regressions developed from the Registry of Cytotoxicity (RC) database.

NICEATM, in coordination with the ICCVAM Acute Toxicity Working Group (ATWG) and ICCVAM, prepared a comprehensive draft background review document (BRD) to describe results and analyses generated from the study. On March 21, 2006, the availability of the draft BRD was announced in an *FR* notice (Vol. 71, No. 54, pp. 14229-14231; available at http://iccvam.niehs.nih.gov/methods/invitro.htm). An international independent Peer Review Panel (hereafter, Panel) convened by ICCVAM on May 23, 2006, reviewed the BRD, evaluated the extent that the BRD addressed established validation and acceptance criteria, and provided comment on the draft ICCVAM recommendations on the use of these test methods, future studies, draft test method protocols, and draft performance standards. On

July 11, 2006, an *FR* notice (Vol. 71, No. 132, pp. 39122-39123; available at http://iccvam.niehs.nih.gov/methods/invitro.htm) announced the availability of the *Peer Review Panel Report: The Use of In Vitro Basal Cytotoxicity Test Methods for Estimating Starting Doses for Acute Oral Systemic Toxicity Testing.* The Panel Report indicated that the information presented in the draft BRD was generally sufficient for its purpose. The Panel concluded that the applicable validation criteria were adequately addressed for use of these *in vitro* test methods in a weight-of-evidence approach to determine starting doses for acute oral toxicity tests.

The accomplishments of the validation study included standardization and optimization of the two NRU protocols that were evaluated and improvement of the LD_{50} database for the 72 reference substances after review of the literature values. The IC_{50} results obtained using the protocols showed that the IC_{50} values in the RC could generally be reproducing with a single cell type and *in vitro* cytotoxicity endpoint. Although the validation study improved the *in vivo* LD_{50} data for the reference chemicals by evaluating LD_{50} values from the scientific literature, IC_{50}[2] -LD_{50} regressions calculated using the validation study data were not different from those calculated using RC data. The validation study also characterized the reproducibility of the NRU test methods and estimated the animal savings that would occur when they are used to determine starting doses for the Up-and-Down Procedure (UDP) (OECD 2001a; EPA 2002a) and the Acute Toxic Class (ATC) method (OECD 2001b).

Accuracy and Reliability

The NICEATM/ECVAM validation study standardized the 3T3 and NHK NRU test methods and improved the LD_{50} database for 72 substances. IC_{50} - LD_{50} regressions were performed for each in vitro NRU test method. The resulting IC_{50} - LD_{50} regressions are consistent with and support continued use of the Registry of Cytotoxicity (RC) database. The RC rat-only millimole regression, which is applicable to substances with known molecular weight, was based on 282 (of 347) RC substances with rat oral LD_{50} data. The RC rat-only data were converted to a weight basis (i.e., mg/kg) to develop the RC rat-only weight regression, which

[2] The IC_{50} is the test substance concentration that produces 50% inhibition of the endpoint measured. The LD_{50} is the median lethal dose.

is applicable to substances without a known molecular weight or to mixtures. The accuracy of the *in vitro* NRU test methods when used with each of the regressions was characterized by determining the proportion of reference substances for which their Globally Harmonized System of Classification and Labelling of Chemicals (GHS; UN 2005) categories (based on rat acute oral LD_{50} data) were correctly predicted.

Using the RC rat-only millimole regression, the 3T3 NRU test method predicted correctly the GHS hazard category of 31% (21/67) of the reference substances successfully tested, while the NHK NRU test method predicted correctly 29% (20/68 reference substances). The accuracy of the 3T3 NRU test method was 69% (46/67 reference substances) for correct category prediction ± 1 category. The corresponding accuracy for the NHK NRU test method was 75% (51/68 reference substances) for correct category prediction ± 1 category.

Using the RC rat-only weight regression, both NRU test methods predicted correctly the GHS hazard category of 31% (21/67 - 3T3; 21/68 - NHK) reference substances successfully tested. The accuracy for the 3T3 NRU test method was 75% (50/67 reference substances) for correct category prediction ± 1 category. The corresponding accuracy for the NHK NRU test method was 75% (51/68 reference substances) for correct category prediction ± 1 category.

Reproducibility was evaluated using the results from the 64 reference substances tested in 3T3 cells and the 68 substances tested in NHK cells that yielded IC_{50} values in all three laboratories (see BRD Section 7 for reliability and reproducibility analyses for the NICEATM/ECVAM validation study). Intra- and inter-laboratory reproducibility of the 3T3 and NHK NRU IC_{50} data were assessed using analysis of variance (ANOVA), coefficient of variation (CV) analysis, comparison of the laboratory-specific IC_{50}-LD_{50} regressions, and comparison of maximum:minimum mean laboratory IC_{50} values.

Results for the positive control (sodium lauryl sulfate [SLS]) IC_{50} values from the 3T3 NRU test method indicated that there were significant differences among laboratories ($p = 0.006$, ANOVA), but not between study phases within laboratories ($p > 0.01$). In addition, interlaboratory CV values were relatively low (2 to 16%). Results for the SLS IC_{50} from the

NHK NRU test method showed significant differences among laboratories ($p < 0.001$) and among study phases within laboratories ($p \leq 0.001$). The use of a different cell culture method at the Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory (FAL) was considered to be responsible for SLS IC_{50} differences among the laboratories in test phases Ia and Ib. Interlaboratory CV values were 39% and 21%, respectively, for phases Ia and Ib, and 31% and 8%, respectively, for phases II and III. The linear regression analyses of the SLS IC_{50} over time (within each laboratory) for both test methods indicated that IC_{50} values generated over the duration of the validation study were stable.

ANOVA analyses for the reference substances showed significant laboratory differences for 23 substances with the 3T3 NRU test method, but only for six substances with the NHK NRU test method (see BRD Tables 7-4 and 7-6). Mean intralaboratory CV values were 26% for both test methods, but the NHK NRU test method had a lower mean interlaboratory CV (28% vs. 47%). (See BRD Tables 7-3 and 7-5 for intra- and inter-laboratory CVs and maximum:minimum ratios.) The maximum:minimum mean laboratory IC_{50} ratios for the 3T3 NRU test method ranged from 1.1 to 21.6, with 52% (33/64) of the reference substances having ratios between 1.5 and 2.5. The maximum:minimum mean laboratory IC_{50} ratios for the NHK NRU test method ranged from 1.0 to 107.6, with 74% (50/68) reference substances having ratios between 1.5 and 2.5. Thus, overall, reproducibility was generally better with the NHK NRU test method.

Animal Welfare

The NICEATM/ECVAM validation study used computer models to simulate the testing of the reference substances in the UDP (OECD 2001a; EPA 2002a) and the ATC method (OECD 2001b), using either the default starting dose (175 mg/kg for the UDP, 300 mg/kg for the ATC) or the starting dose predicted by the 3T3 and NHK NRU test methods (see BRD Section 10 for simulation modeling and analyses for the study). The simulations were used to estimate, per substance, the number of animals that would be used and their associated survival rate. The modeling was performed using five different dose-mortality slopes[3] (i.e., 8.3, 4.0, 2.0, 0.8, and 0.5) because slope information was not available for many of the

[3] The dose-mortality slope is the slope of the dose-response curve for mortality.

reference substances. Both RC rat-only regressions were used to determine starting doses from IC_{50} data obtained using either the 3T3 or NHK NRU test methods. In principle, animal savings with the Fixed Dose Procedure (FDP; OECD 2001c) could be estimated even though death is not the primary endpoint, but the validation study did not include this analysis.

Computer simulation of the UDP testing showed that, for the substances with rat acute oral LD_{50} reference data tested in the validation study (67 substances for 3T3, 68 substances for NHK) an average of 0.49 animals (6.2%) to 0.66 animals (7.0%) would be saved. No animal savings were predicted for reference substances with $50 < LD_{50} \leq 300$ mg/kg, which is where the default dose of 175 mg/kg occurs. The highest animal savings were predicted for substances with $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$ mg/kg for both NRU test methods (1.28 [11.9%] to 1.65 animals [16.7%]) because limit testing calls for fewer animals than the main test. The greatest animal savings were observed for substances in these categories because the limit test, which would be used for such substances, uses fewer animals that the main test. Although using the 3T3 and NHK NRU IC_{50} values to estimate starting doses for the simulated UDP deceased the number of animals used, it did not change the number of animals that would be expected to be euthanized or die.

Computer simulation of ATC testing showed that, for the substances tested in the validation study, NRU test methods resulted in an average savings of 0.51 animals (4.8%) to 1.09 animals (10.2%) per test. No animal savings were predicted for substances with $300 < LD_{50} \leq 2000$ mg/kg, which is where the default dose of 300 mg/kg would have been used. Mean animal savings for substances with $2000 < LD_{50} \leq 5000$ mg/kg ranged from -0.03 animals (-0.03%) to 0.11 animals (0.9%) for the RC rat-only millimole regression and from 0.53 animals (4.7%) to 2.43 animals (20.5%) for the RC rat-only weight regression. For both regressions evaluated, mean animal savings for substances with $LD_{50} > 5000$ mg/kg ranged from 2.03 animals (17.1%) to 3.33 animals (27.7%). The greatest reduction in animal use occurs for substances in this category because the limit test uses fewer animals than the main test.

The magnitude of animal savings did not correlate with the accuracy of GHS categorization yielded by the NRU-predicted LD_{50} values (using the *in vitro* NRU IC_{50} values in the IC_{50}-LD_{50} regressions) or with the accuracy of GHS category outcomes because the accuracy and animals savings analyses used different standards for comparison (see BRD Section 10.4).

The use of the IC_{50}-based starting doses did not significantly alter the GHS category outcomes of the simulated UDP (based on LD_{50} outcome) or ATC when compared with the outcomes based on the default starting dose. The concordance for GHS acute oral toxicity category for the IC_{50}-based starting dose with the default starting dose was 97 to 99% for both *in vitro* NRU methods and IC_{50}-LD_{50} regressions evaluated.

ICCVAM Test Method Recommendations for Uses and Future Studies

ICCVAM recommends that, while the 3T3 and NHK NRU test methods are not sufficiently accurate to predict acute oral toxicity for the purpose of regulatory hazard classification, they may be used in a weight-of-evidence approach to determine the starting dose for the current acute oral toxicity protocols (i.e., the UDP, the ATC) (see **Sections 2.6** and **2.7**). Therefore, ICCVAM recommends that use of the 3T3 and NHK NRU test methods be considered[4] before an acute oral toxicity test is initiated. Use of the NRU data with the RC rat-only millimole regression generally underpredicted toxicity for substances in the highest toxicity (i.e., lowest LD_{50}) categories and overpredicted toxicity for substances in the lowest toxicity (i.e., highest LD_{50}) categories (see BRD Table 6-5). Although substances at the very low and high ends of the toxicity range were poorly predicted, those in the middle range (i.e., $300 < LD_{50} \leq 2000$ mg/kg) were predicted better. Substances with specific toxic mechanisms, such as neurotoxicity or cardiotoxicity are not expected to be cytotoxic. Many highly toxic substances have specific mechanisms (e.g., receptor-mediated effects) that cytotoxicity systems would not be expected to detect. Such substances are likely to be underpredicted by these *in vitro* basal cytotoxicity test methods and these methods may not be appropriate for estimating starting doses for such substances.

[4] The 3T3 NRU test method is recommended for general use because it is less labor intensive and less expensive to conduct than the NHK NRU test method.

The regression formula used to determine starting doses for test substances with known molecular weights should be the RC rat-only millimole regression (log LD_{50} mmol/kg = 0.439 log IC_{50} mM + 0.621) with IC_{50} values in mmol/L and LD_{50} values in mmol/kg. The regression formula for mixtures or test substances with unknown molecular weights should be the RC rat-only weight regression, with IC_{50} values in μg/mL and LD_{50} values in mg/kg (i.e., log LD_{50} mg/kg = 0.372 log IC_{50} μg/mL + 2.024).

For future studies to advance the use of alternative test methods for predicting acute oral toxicity, additional data should be collected using the 3T3 NRU basal cytotoxicity test method to evaluate its usefulness for predicting the rodent acute oral toxicity of chemical mixtures. To supplement the high quality database developed in the NICEATM/ECVAM validation study, additional high quality *in vitro* basal cytotoxicity data should be collected when rat acute oral toxicity testing is conducted. Such data can be used in periodic evaluations to further characterize the usefulness and limitations of using *in vitro* cytotoxicity data as part of a weight-of-evidence approach to estimate starting doses.

Additional efforts should be made to identify *in vitro* tests and other methods necessary to achieve accurate acute oral hazard classification without the use of animals. The *in vivo* database of LD_{50} values for reference substances used in this validation study should be used to evaluate the utility of other non-animal approaches to estimate starting doses for acute oral toxicity tests. An expanded list of reference substances with rat acute oral LD_{50} values substantiated by high quality *in vivo* data (including data currently held by industry) should be developed for use in future *in vitro* test method development and validation studies. To support the development of mechanism-based *in vitro* methods, standardized procedures to collect *in vivo* measurements and observations pertinent to an understanding of the mechanisms of lethality should be included in future rat acute oral toxicity studies.

Performance Standards

The purpose of performance standards is to communicate the basis by which adequately validated new proprietary (e.g., copyrighted, trademarked, registered) and nonproprietary test

methods have been determined to have sufficient accuracy and reliability for specific testing purposes (see **Section 3**). The three elements of performance standards are:

- Essential test method components (i.e., structural, functional, and procedural elements of a validated test method that a proposed, mechanistically and functionally similar test method should adhere to)
- A minimum list of reference chemicals for assessing the accuracy and reliability of the proposed test method
- The accuracy and reliability values that should be achieved by the proposed test method using the minimum list of reference chemicals.

The performance of other *in vitro* basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD_{50}, respectively) should be demonstrated to meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods as determined in the NICEATM/ECVAM validation study.

To demonstrate technical proficiency with the validated 3T3 or NHK NRU test methods, ICCVAM recommends that the user evaluate his/her ability to calculate IC_{50} values for a minimum of two unclassified chemicals and two from each from the five GHS hazard categories (i.e., at least 12 of the 30 reference substances listed in **Table 3-1**). The resulting IC_{50} values should be within 2.5 standard deviations of the reported IC_{50} values in the table. Intralaboratory CV values for the IC_{50} of the reference substances should not exceed 129% for the NHK NRU test method or 98% for the 3T3 NRU test method and the mean CV for the substances tested should not exceed 26% for the NHK NRU test method or 27% for the 3T3 NRU test method.

Before using a candidate *in vitro* basal cytotoxicity test method to predict starting doses, the correlation between the *in vitro* and the *in vivo* test methods must be established quantitatively. This can be accomplished by using the new test method to test a subset of the 30 reference substances that cover all six hazard classification categories (i.e., the entire range of acute oral toxicity) and that produce the same regression formula as the total

database. After testing, the IC_{50} data should be used to calculate a linear regression formula (least square method) for the reference substances using the corresponding LD_{50} values (provided in **Table 3-1)**. The resulting regression is compared against a regression calculated using the IC_{50} and LD_{50} values from the NICEATM/ECVAM validation study (in **Table 3-1**). If a comparison of slope and intercept indicates that the two regressions are not statistically significantly different (at $p < 0.05$), then the test is considered suitable to generate IC_{50} data to use with the recommended regression formula for estimating starting doses for acute oral toxicity/lethality tests.

Candidate basal cytotoxicity test methods should achieve the overall accuracy of the 3T3 NRU test method for correctly predicting the GHS acute oral toxicity classification category of the 30 reference substances, which was 33% for the RC rat-only millimole regression and 30% for the RC rat-only weight regression.

1.0 INTRODUCTION

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) is charged by the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 [2000]; available at http://iccvam.niehs.nih.gov/about/PL106545.pdf) with evaluating the scientific validity of new, revised, and alternative toxicological test methods applicable to U.S. Federal agency safety testing requirements. Following such evaluations, ICCVAM is required to provide recommendations to U.S. Federal agencies regarding the usefulness and limitations of such methods.

1.1 Evaluation of the Use of *In Vitro* Cytotoxicity Test Methods to Estimate Acute Oral Toxicity

ICCVAM initiated a review of the validation status of *in vitro* methods for estimating acute oral toxicity in 1999, in response to a request from the U.S. Environmental Protection Agency (EPA) Office of Pesticides, Prevention, and Toxic Substances. This request was based on recently published studies that showed a correlation between *in vitro* cytotoxicity and *in vivo* acute toxicity. In October of 2000, the National Toxicology Program (NTP), the National Institute of Environmental Health Sciences (NIEHS), and the EPA sponsored the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity, which was announced in the *Federal Register* [(*FR*); Vol. 65, No. 184, pp. 57203-57205; available at http://iccvam.niehs.nih.gov/methods/invidocs/6557203.htm). Invited scientific experts and ICCVAM agency scientists were assigned to one of four Breakout Groups and prepared recommendations on the following:

- *In Vitro* Screening Methods for Assessing Acute Toxicity
- *In Vitro* Methods for Toxicokinetic Determinations
- *In Vitro* Methods for Predicting Organ Specific Toxicity
- Chemical Data Sets for Validation of *In Vitro* Acute Toxicity Test Methods

Workshop participants concluded that none of the proposed *in vitro* methods reviewed had been formally evaluated for reliability and relevance, and that their usefulness and limitations for generating information to meet regulatory requirements for acute toxicity testing had not been adequately assessed. However, an *in vitro* approach previously proposed by ZEBET

(the German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments) was recommended by workshop participants as a high priority for rapid adoption so that data could be generated to establish its usefulness with a large number of chemicals (ICCVAM 2001a). The proposal was to use *in vitro* cytotoxicity data to estimate starting doses for in vivo acute toxicity studies. Since a correlation between IC_{50} and LD_{50} values had been determined based on retrospective literature reviews, such a strategy might. reduce the use of animals for acute oral toxicity tests by identifying a starting dose closer to the LD_{50}[5]. To provide sample *in vitro* cytotoxicity protocols and instructions for using *in vitro* data to predict starting doses for acute rodent oral toxicity tests, the *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (ICCVAM 2001b) was prepared by workshop participants with the assistance of ICCVAM and NICEATM.

1.2 Evaluation of the Use of *In Vitro* Cytotoxicity Test Methods to Estimate Starting Doses for Acute Oral Toxicity Tests

ICCVAM agreed with workshop participants that *in vitro* basal cytotoxicity test methods should have a high priority for validation studies. Therefore, the NTP Center for the Evaluation of Alternative Toxicological Methods (NICEATM) collaborated with the European Centre for the Validation of Alternative Methods (ECVAM), a component of the European Commission's Joint Research Centre, to further characterize the usefulness of *in vitro* cytotoxicity assays as predictors of starting doses for acute oral systemic toxicity assays. NICEATM and ECVAM designed a multi-laboratory validation study using 72 reference substances to evaluate the performance of two standardized *in vitro* neutral red uptake (NRU) test methods, based on the ZEBET approach using the Registry of Cytotoxicity (RC) millimole regression model. The objectives for the validation study were to:

- Further standardize and optimize the *in vitro* NRU cytotoxicity protocols using BALB/c 3T3 mouse fibroblasts (3T3) and normal human epidermal keratinocytes (NHK) to maximize test method reliability (intralaboratory repeatability, intra- and inter-laboratory reproducibility)

[5] LD_{50}: Lethal dose that produces lethality in 50% of test animals

- Assess the accuracy of the two standardized *in vitro* 3T3 and NHK NRU basal cytotoxicity test methods for estimating rodent oral LD_{50} values across the five United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS; UN 2005) categories of acute oral toxicity, as well as unclassified toxicities
- Estimate the reduction and refinement in animal use achievable from using the *in vitro* 3T3 and NHK NRU test methods to identify starting doses for *in vivo* acute oral toxicity tests, assuming that no other information were available
- Develop high quality *in vivo* acute oral lethality and *in vitro* NRU cytotoxicity databases that can be used to support the investigation of other *in vitro* test methods necessary to improve the prediction of *in vivo* acute oral lethality

The validation study proceeded in four phases so that the Study Management Team (SMT) could evaluate the reproducibility of results after each phase and refine the protocols, if necessary, before proceeding to the next phase. Three laboratories participated in testing the 72 reference substances using the 3T3 and NHK NRU test methods, beginning in August 2002 and ending in January 2005:

- The U.S. Army Edgewood Chemical Biological Center, Edgewood, MD (ECBC)
- Fund for the Replacement of Animals in Medical Experiments Alternatives Laboratory, Nottingham, UK (FAL)
- The Institute for *In Vitro* Sciences, Gaithersburg, MD (IIVS)

BioReliance Corporation (Rockville, MD) procured and distributed the coded reference substances and performed solubility tests prior to their distribution to the testing laboratories, but did not perform any of the *in vitro* tests.

NICEATM, in coordination with the ICCVAM Acute Toxicity Working Group (ATWG) and ICCVAM, prepared a comprehensive draft background review document (BRD) to summarize the procedures and results generated from the validation study. On March 21, 2006, the availability of the draft BRD was announced in an *FR* notice (Vol. 71, No. 54, pp.

14229-14231; available at http://iccvam.niehs.nih.gov/methods/invitro.htm). The BRD was made available to the public in electronic format on the ICCVAM/NICEATM website (available at http://iccvam.niehs.gov) and in print upon request to NICEATM.

1.3 Peer Review of the NICEATM/ECVAM Validation Study

An international independent Peer Review Panel (hereafter, Panel) convened by ICCVAM on May 23, 2006, reviewed the BRD, evaluated the extent that the BRD addressed established validation and acceptance criteria, and provided comment on the draft ICCVAM recommendations on the use of these test methods, future studies, draft test method protocols, and draft performance standards. Comments from the public and scientific community were provided to the Panel and made available on the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov/methods/invidocs/brdcomm.htm). On July 11, 2006, an *FR* notice (Vol. 71, No. 132, pp. 39122-39123; available at http://iccvam.niehs.nih.gov/methods/invitro.htm) announced the availability of the *Peer Review Panel Report: The Use of In Vitro Basal Cytotoxicity Test Methods for Estimating Starting Doses for Acute Oral Systemic Toxicity Testing.* The Panel report (see **Appendix A**) indicated that the information presented in the draft BRD was generally sufficient for its purpose. The Panel concluded that the objectives of the validation study were appropriate, and agreed that the applicable validation criteria were adequately addressed for use of these *in vitro* test methods in a weight-of-evidence approach to determine starting doses for acute oral toxicity tests.

Regarding the draft ICCVAM recommendations for test method uses, the Panel agreed that:

- Neither of the NRU test methods can be used as alternatives for the *in vivo* acute oral toxicity test for the purposes of hazard classification.
- The *in vitro* NRU test methods may be useful in a weight-of-evidence approach to determine the starting dose for acute oral *in vivo* toxicity protocols.
- The NRU test methods should be considered before animals are used.

- The RC rat-only regression should be used to estimate the LD_{50} from IC_{50}[6] data. When the molecular weight of a test substance is known, the molar regression should be used; however, a regression based on weight rather than molar units should be used when the molecular weight of the test substance is unknown.
- Other *in vitro* basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD_{50} value, respectively) should meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods.
- The 3T3 NRU test method, based on relative ease of performance and cost, should be recommended for general use, but cautioned that one test method should not be preferred over the other.
- The NRU test methods are appropriate for substances that interfere with energy utilization or alkylation of proteins and other macromolecules

Regarding the draft ICCVAM recommendations for future studies, the Panel agreed that:

- Additional data for the 3T3 NRU test method should be collected to evaluate its usefulness for predicting starting doses with chemical mixtures.
- High quality comparative *in vitro* basal cytotoxicity data should be collected in tandem with *in vivo* rat acute oral toxicity test results to further evaluate the use of these test methods for predicting the starting doses for acute oral toxicity tests.
- Additional *in vitro* tests and other methods necessary to achieve accurate acute oral hazard classification should be investigated.
- The *in vivo* database of reference substances used in the validation study should be used to evaluate the utility of other non-animal approaches to estimate starting doses for rat acute oral toxicity tests.

[6] IC_{50}: Test substance concentration producing 50% inhibition of the endpoint measured.

- Standardized procedures to collect information pertinent to an understanding of the mechanisms of lethality should be included, to the extent possible, in future rat acute oral toxicity studies.
- An expanded list of reference substances with estimated rat LD_{50} values substantiated by high quality *in vivo* data should be developed for use in future *in vitro* test development and validation.

The draft BRD, the Panel report, and all associated public comments were made available to the Scientific Advisory Committee on Alternative Toxicological Methods (SACATM) for their consideration. The SACATM endorsed the Panel Report. ICCVAM and the ATWG then considered the Panel Report, all public comments, and the comments of SACATM in preparing the final test method recommendations that are provided in this report. This report will be made available to the public and provided to U.S. Federal agencies for consideration, in accordance with the ICCVAM Authorization Act of 2000 (42 U.S.C. § 2851-2, 2851-5 [2000]; available at http://iccvam.niehs.nih.gov/about/PL106545.pdf). The final BRD, revised in response to the Panel and ATWG comments, will also be provided as background information and technical support for this report. Agencies with applicable testing regulations and guidelines (**Appendix B**) are required by law to respond to ICCVAM within 180 days of receiving the ICCVAM recommendations. These responses will be made available to the public on the ICCVAM website (http://iccvam.niehs.nih.gov) as they are received.

1.4 Report Organization

Section 1.0 of this report provides the background of the NICEATM/ECVAM validation study for the use of *in vitro* cytotoxicity test methods to predict starting doses for acute oral toxicity test methods and this resulting ICCVAM test method evaluation report. **Section 2.0** describes the NRU protocols evaluated in the validation study, the reference substances tested, and the accuracy and reliability results from the validation study. Also included are ICCVAM's recommendations for test method uses and future studies, which were finalized after consideration of the Panel Report, public comments, and the comments of SACATM, and were based on the results of the validation study. The recommendations for future studies are intended to advance the use of alternative methods for the prediction of acute toxicity. **Section 3.0** provides recommended performance standards for application to future test

methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect. The three elements of performance standards are essential test method components (i.e., structural, functional, and procedural elements of a validated test method that a proposed, mechanistically and functionally similar test method should adhere to), a minimum list of reference chemicals for assessing the accuracy and reliability of the proposed test method, and the accuracy and reliability values that should be achieved by the proposed test method using the minimum list of reference chemicals.

2.0 ICCVAM RECOMMENDATIONS FOR *IN VITRO* NRU BASAL CYTOTOXICITY TEST METHODS

The following technical summary provides a synopsis of the performance analysis described in the BRD which indicates the current validation status of the *in vitro* 3T3 and NHK NRU basal cytotoxicity test methods, including what is known about their reliability and accuracy, the scope of the substances tested, and standardized protocols. These results form the basis for the ICCVAM Recommendations for test method uses and future studies that are presented at the end of this section.

2.1 Test Method Description

The NRU cytotoxicity assay procedure is based on the ability of viable cells to incorporate and bind neutral red (NR), a supravital dye. NR is a weakly cationic dye that readily diffuses through the plasma membrane and concentrates in lysosomes where it electrostatically binds to the anionic lysosomal matrix. Toxicants can alter the cell surface or the lysosomal membrane to cause lysosomal fragility and other adverse changes that gradually become irreversible. Thus, cell death and/or inhibition of cell growth decreases the amount of NR retained by the culture. Healthy proliferating mammalian cells, when properly maintained in culture, continuously divide and multiply over time. A toxic substance, regardless of site or mechanism of action, will interfere with this process and result in a reduction of the growth rate as reflected by cell number. Cytotoxicity is expressed as a concentration dependent reduction of the uptake of NR after substance exposure to the cells, thus providing a sensitive, integrated signal of both cell integrity and growth inhibition.

2.1.1 General Test Method Procedures

3T3 and NHK cell cultures are grown in 96-well microtiter plates and exposed to a reference substance and/or positive control (PC). After the predetermined incubation time, the reference substance and PC are removed and NR solution is applied to the cells. The cells are incubated again, the excess NR solution is removed, and NR is eluted from the cells. The NRU is determined by using a microtiter plate reader/spectrophotometer to measure the optical density (OD; at a wavelength of 540 ± 10 nm) of the eluted NR dye in the 96-well plate. A calculation of cell viability expressed as NRU is made for each concentration of a reference substance and PC by using the mean NRU OD of six replicate values (minimum of

four acceptable replicate wells) per test concentration. The cell viability OD value is compared with the mean NRU OD of all vehicle control (VC) values (provided VC values have met the VC acceptance criteria). Relative cell viability is then expressed as percentage of untreated VC.

2.1.2 Protocol Similarities and Differences for the 3T3 and NHK NRU Test Methods

A number of protocol procedures and conditions are common to both the 3T3 and NHK NRU test methods (see **Appendices C1** and **C2** for specific protocols for the test methods). Both NRU test methods use the same solvents to dissolve reference substances and the PC, the same culture conditions, the same 96-well plate format, and the same duration of exposure, and both employ the use of a range finder test before performing the definitive (main) test. In addition, both NRU test methods follow identical NRU procedures and calculate cell viability and the IC_{50} using the same procedures.

There are three differences between the protocols for the 3T3 and NHK NRU test methods. The first is the use of newborn calf serum in the 3T3 cell culture medium. The NHK cells require a keratinocyte-specific serum-free medium. The second is that the 3T3 cells require less time (approximately 24 hours) to reach appropriate the confluence for testing than the NHK cells (approximately 24 to 72 hours). The third difference is the application and volume of test substance. For the 3T3 NRU test method, all culture medium is removed from the 3T3 cells and 50 µL/well of medium with substance is added immediately. For the NHK cells, 125 µL/well of medium with test substance is added to the 125 µL/well of medium already on the cells.

2.2 Reference Substances

Seventy-two reference substances were selected for testing in the NICEATM/ECVAM validation study. These substances were selected to represent: (1) the complete range of *in vivo* acute oral LD_{50} values; (2) the types of substances regulated by the various regulatory authorities; and (3) those with human toxicity data and/or human exposure potential. To insure that the complete range of toxicity was covered, the GHS (UN 2005) was used to select 12 substances for each of five acute oral toxicity categories and 12 unclassified substances. The set of selected reference substances had the following characteristics:

- Thirty-five percent (27/77) were pharmaceuticals, 22% (17/77) were pesticides, 10% (8/77) were solvents, and 6% (5/77) were food additives. The remaining substances were used for a variety of manufacturing and consumer products. The number of assigned uses (77) is greater than the number of selected substances because some of the substances have more than one use.
- Relevance of the substances to human exposure is indicated by the fact that 58% (42/72) were included in the Multicentre Evaluation of *In Vitro* Cytotoxicity (MEIC) study, 24% (17/72) of which were included also in the Evaluation-guided Development of New *In Vitro* Tests (EDIT) program; 64% (46/72) had human exposures reported by the Toxic Exposure Surveillance System (TESS); 71% (51/72) had been evaluated by NTP; and 25% (18/72) were on the EPA High Production Volume (HPV) list.
- Eighty-one percent (58/72) of the substances were in the RC database[7]; 38% (22/58) of which were outliers with respect to the RC millimole regression (log LD_{50} mmol/kg = 0.435 x log IC_{50} mM + 0.625). The RC millimole regression underpredicted the toxicity of 77% (17/22) of the outliers and overpredicted the toxicity of 23% (5/22).
- Seventy-nine percent (57/72) were organic compounds and 21% (15/72) were inorganic. The most commonly represented classes of organic compounds were heterocyclics (25%, 14/57), carboxylic acids (25%, 14/56), and alcohols (18%, 10/57).
- Twenty-six percent (19/72) were known to have active metabolites and three others were expected to have active metabolites based on their chemical structures.
- Many of the substances produced toxicity in more than one organ system. The most common target organs were liver (17 substances) and kidney (15 substances). Other target organs included the nervous system (40 substances)

[7] The RC is a database of acute oral LD_{50} values for rats and mice obtained from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for 347 chemicals with known molecular weights (Halle 1998, 2003).

and cardiovascular system (10 substances). No target organ information was available for one substance (gibberellic acid).

2.3 Test Method Accuracy

The ability of the 3T3 and NHK NRU test methods to correctly predict rodent acute oral systemic toxicity is based on the validity of the *in vivo – in vitro* (i.e., IC_{50}-LD_{50}) regression model. It is the IC_{50}-LD_{50} regression that establishes the relationship between the 3T3 and NHK NRU IC_{50} values and the predicted LD_{50} values that were used to set the starting doses for the computer simulated acute oral toxicity assays performed for the NICEATM/ECVAM validation study.

The validation study tested two regressions for their use with the NRU IC_{50} values to predict LD_{50} values. The first regression – the RC rat-only millimole regression – was calculated using the 282 substances in the RC dataset of 347 substances that had a reported rat oral LD_{50} value (65 substances had mouse-only LD_{50} values). The LD_{50} data for the regression were limited to one species to decrease the variability in LD_{50} values produced by combining the data of two species. Rats were selected because they are the preferred species for most acute oral toxicity testing (i.e., UDP, ATC, and FDP; EPA 2002b; OECD 2001a; OECD 2001d). The second regression – the RC rat-only weight regression was a transformation of the RC rat-only millimole regression to weight units (mg/kg body weight for LD_{50} and µg/mL for IC_{50}) in order to make the regression applicable to the testing of mixtures and substances without a known molecular weight.

The ability of the 3T3 and NHK NRU IC_{50} data to correctly predict rat acute oral LD_{50} values, based on using the RC rat-only millimole regression and the RC rat-only weight regression, was evaluated by determining the extent to which the appropriate GHS acute oral toxicity category was identified for each reference substance. This approach permits an assessment of accuracy specific to each GHS hazard classification category.

Tables 2-1 and **2-2**, which are divided into upper and lower sub-tables, provide accuracy data for the 3T3 and NHK NRU test methods, respectively. For each part, the toxicity categories corresponding to the reference rat acute oral LD_{50} data are provided in rows that are labeled

on the far left side of the table. The toxicity categories predicted by the *in vitro* NRU assays and the associated regressions are provided in columns that are labeled across the top of each part (i.e., 3T3 or NHK NRU-predicted toxicity category) of the table. The numbers at the intersections of the reference rat oral LD_{50} rows and 3T3 or NHK NRU-predicted toxicity category columns are the numbers of substances with *in vitro* category predictions that correspond to the various *in vivo* categories. The right sides of the tables also provide summaries containing, for each rat acute oral toxicity category and for the total number of substances evaluated:

- The number of substances
- The accuracy of the 3T3 or NHK NRU prediction
- The percentage of substances for which toxicity has been overpredicted and underpredicted by the *in vitro* NRU test methods.

In each of the 3T3 and NHK sections of the table, a summary of predictivity is also provided for each predicted toxicity category along with the percentage of substances with category (i.e., toxicity) underpredicted and overpredicted.

Table 2-1 shows the concordance of the observed (i.e., the rat acute oral LD_{50}) and predicted GHS acute oral toxicity categories (UN 2005) for each *in vitro* NRU cytotoxicity test method using the geometric mean IC_{50} values (of the three validation study laboratories) in the RC rat-only millimole regression, $\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621. Accuracy is the agreement of the *in vitro* NRU category predictions with those based on the rat acute oral LD_{50} reference values.

Table 2-1 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Millimole Regression[1]

Reference Rat Oral LD_{50}[2] (mg/kg)	3T3 NRU-Predicted GHS Category (mg/kg)						Total	Accuracy	Toxicity Over-predicted	Toxicity Under-predicted
	LD_{50} <5	5< LD_{50} ≤50	50 < LD_{50} ≤300	300 < LD_{50} ≤2000	2000 < LD_{50} ≤5000	LD_{50} >5000				
LD_{50} < 5	0	2	0	4	0	0	6[3]	0%	0%	100%
5< LD_{50} ≤50	0	1	6	3	1	0	11[4]	9%	0%	91%
50 < LD_{50} ≤300	0	0	5	7	0	0	12	42%	0%	58%
300 < LD_{50} ≤2000	0	1	2	13	0	0	16	81%	19%	0%
2000 < LD_{50} ≤5000	0	0	0	10	0	0	10[5]	0%	100%	0%
LD_{50} >5000	0	0	0	8	2	2	12[6,7]	17%	83%	0%
Total	0	4	13	45	3	2	67	31%	34%	34%
Predictivity	0%	25%	38%	29%	0%	100%				
Category Overpredicted	0%	50%	46%	31%	33%	0%				
Category Underpredicted	0%	25%	15%	40%	67%	0%				
Reference Rat Oral LD_{50}[2]	**NHK NRU-Predicted Toxicity Category (mg/kg)**						**Total**	**Accuracy**	**Toxicity Over-predicted**	**Toxicity Under-predicted**
	LD_{50} <5	**5< LD_{50} ≤50**	**50 < LD_{50} ≤300**	**300 < LD_{50} ≤2000**	**2000 < LD_{50} ≤5000**	**LD_{50} >5000**				
LD_{50} <5	0	1	2	3	0	0	6[3]	0%	0%	100%
5< LD_{50} ≤50	0	2	5	3	1	0	11[4]	18%	0%	82%
50 < LD_{50} ≤300	0	1	6	5	0	0	12	50%	8%	42%
300 < LD_{50} ≤2000	0	1	2	12	1	0	16	75%	19%	6%
2000 < LD_{50} ≤5000	0	0	0	10	0	0	10[5]	0%	100%	0%
LD_{50} >5000	0	0	0	7	6	0	13[7]	0%	100%	0%
Total	0	5	15	40	8	0	68	29%	40%	31%
Predictivity	0%	40%	40%	30%	0%	0%				
Category Overpredicted	0%	20%	47%	28%	25%	0%				
Category Underpredicted	0%	40%	13%	43%	75%	0%				

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); 3T3=BALB/c 3T3 fibroblasts; NHK=normal human keratinocytes; NRU=neutral red uptake; RC=Registry of Cytotoxicity.

[1]The RC rat-only millimole regression is log LD_{50} (mmol/kg) = log IC_{50} (mM) x 0.439 + 0.621. Numbers in table represent numbers of substances.

[2]Reference rat oral LD_{50} values in mg/kg from (see BRD Table 4-2**)**

[3]Epinephrine bitartrate excluded because no rat reference oral LD_{50} was identified (BRD Table 4-2**)**

[4]Colchine excluded because no rat LD_{50} was identified (BRD Table 4-2)

[5]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[6]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[7]Propylparaben excluded because no rat LD_{50} was identified (see BRD Table 4-2).

Note: BRD Table 4-2 can be found at http://iccvam.niehs.nih.gov/methods/invitro.htm

The overall accuracy of the 3T3 NRU test method for correctly predicting GHS acute oral toxicity classification category using the RC rat-only millimole regression was 31% (21/67 substances). Rat acute oral toxicity was overpredicted for 34% (23) and underpredicted for 34% (23) of the 67 substances. For this analysis, in terms of each GHS acute oral toxicity classification category:

- Zero (0%) of six substances with LD_{50} <5 mg/kg was correctly predicted
- One (9%) of 11 substances in the 5 < LD_{50} ≤50 mg/kg category was correctly predicted
- Five (42%) of 12 substances in the 50 < LD_{50} ≤300 mg/kg category were correctly predicted
- Thirteen (81%) of 16 substances in the 300 < LD_{50} ≤2000 mg/kg category were correctly predicted; however, this toxicity category was also predicted for 32 other substances (71%; 32/45) that did not match this category *in vivo*. Thus, the predictivity for this category was 29% (13/45 substances predicted for this category matched the *in vivo* category).
- None (0%) of the 10 substances in the 2000 < LD_{50} ≤5000 mg/kg category were correctly predicted
- Two (17%) of the 12 substances with LD_{50} >5000 mg/kg were correctly predicted

The overall accuracy of the NHK NRU test method for correctly predicting the GHS acute oral toxicity classification, when the prediction was based on the RC rat-only millimole regression, was 29% (20/68 substances). Toxicity was overpredicted for 40% (27) and underpredicted for 31% (21) of the 68 substances. The pattern of concordance between *in vitro* and *in vivo* results for the NHK NRU test method with the RC rat-only millimole regression was similar to that for the 3T3 NRU test method with the exception that the toxicity of all substances with LD_{50} >50000 mg/kg were not correctly predicted. For this analysis, in terms of each GHS acute oral toxicity classification category:

- Zero (0%) of six substances with LD_{50} <5 mg/kg were correctly predicted
- Two (18%) of 11 substances in the 5 < LD_{50} ≤50 mg/kg category were correctly predicted

- Six (50%) of 12 substances in the $50 < LD_{50} \leq 300$ mg/kg categories were correctly predicted
- 12 (75%) of 16 substances in the $300 < LD_{50} \leq 2000$ mg/kg category were correctly predicted; however, this toxicity category was also predicted for 28 (70%; 28/40) other substances with *in vivo* data that did not match the category. Thus, the predictivity for this category was 30% (12/40).
- Zero (0%) of 10 substances in the $2000 < LD_{50} \leq 5000$ mg/kg category were correctly predicted
- None (0%) of 13 substances with $LD_{50} > 5000$ mg/kg were correctly predicted.

Table 2-2 shows the concordance of the observed and predicted GHS acute oral toxicity categories for each *in vitro* NRU test method using the geometric mean IC_{50} values (of the three validation study laboratories) and the RC rat-only weight regression. The regression formula for the RC rat-only weight regression is $\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (µg/mL) + 2.024.

Table 2-2 Prediction of GHS Acute Oral Toxicity Category by the 3T3 and NHK NRU Test Methods and the RC Rat-Only Weight Regression[1]

Reference Rat Oral LD_{50}[2] (mg/kg)	3T3 NRU-Predicted Toxicity Category (mg/kg)						Total	Accuracy	Toxicity Over-predicted	Toxicity Under-predicted
	LD_{50} <5	5 < LD_{50} ≤50	50 < LD_{50} ≤300	300 < LD_{50} ≤2000	2000 < LD_{50} ≤5000	LD_{50} >5000				
LD_{50} <5	0	0	2	4	0	0	6[3]	0%	0%	100%
5 < LD_{50} ≤50	0	1	5	5	0	0	11[4]	9%	0%	91%
50 < LD_{50} ≤300	0	0	4	8	0	0	12	33%	0%	67%
300 < LD_{50} ≤2000	0	1	3	12	0	0	16	75%	25%	0%
2000 < LD_{50} ≤5000	0	0	0	6	4	0	10[5]	40%	60%	0%
LD_{50} >5000	0	0	0	5	7	0	12[6,7]	0%	100%	0%
Total	0	2	14	40	11	0	67	31%	33%	36%
Predictivity	0%	50%	29%	30%	36%	0%				
Category Overpredicted	0%	0%	50%	43%	0%	0%				
Category Underpredicted	0%	50%	21%	28%	64%	0%				
Reference Rat Oral LD_{50}[2] (mg/kg)	**NHK NRU-Predicted Toxicity Category (mg/kg)**									
	LD_{50} <5	**5 < LD_{50} ≤50**	**50 < LD_{50} ≤300**	**300 < LD_{50} ≤2000**	**2000 < LD_{50} ≤5000**	**LD_{50} >5000**	**Total**	**Accuracy**	**Toxicity Over-predicted**	**Toxicity Under-predicted**
LD_{50} <5	0	1	2	3	0	0	6[3]	0%	0%	100%
5 < LD_{50} ≤50	0	1	5	5	0	0	11[4]	9%	0%	91%
50 < LD_{50} ≤300	0	1	5	6	0	0	12	42%	8%	50%
300 < LD_{50} ≤2000	0	1	2	13	0	0	16	81%	19%	0%
2000 < LD_{50} ≤5000	0	0	0	9	1	0	10[5]	10%	90%	0%
LD_{50} >5000	0	0	0	6	6	1	13[7]	8%	92%	0%
Total	0	4	14	42	7	1	68	31%	37%	32%
Predictivity	0%	25%	36%	31%	14%	100%				
Category Overpredicted	0%	25%	50%	33%	0%	0%				
Category Underpredicted	0%	50%	14%	36%	86%	0%				

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); 3T3=BALB/c 3T3 fibroblasts; NHK=Normal human keratinocytes; NRU=Neutral red uptake; RC=Registry of Cytotoxicity.

[1]The RC rat-only weight regression is log LD_{50} (mg/kg) = log IC_{50} (μg/mL) x 0.372 + 2.024.

[2]Reference rat oral LD_{50} values in mg/kg from BRD Table 4-2.

[3]Epinephrine bitartrate excluded because no rat LD_{50} was identified (see BRD Table 4-2).

[4]Colchine excluded because no rat LD_{50} was identified (see BRD Table 4-2).

[5]Carbon tetrachloride excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[6]Methanol excluded because no laboratory attained sufficient toxicity for the calculation of an IC_{50}.

[7]Propylparaben excluded because no rat LD_{50} was identified (see BRD Table 4-2).

Note: BRD Table 4-2 can be found at http://iccvam.niehs.nih.gov/methods/invitro.htm

The overall accuracy of the 3T3 NRU test method with the RC rat-only weight regression was 31% (21) for the results from 67 substances. The toxicity was overpredicted for 33% (24) and underpredicted for 36% (22) of the 67 substances. For this analysis, in terms of each GHS acute oral toxicity classification category:

- Zero (0%) of six substances with LD_{50} <5 mg/kg were correctly predicted
- One (9%) of 11 substances in the 5 < LD_{50} ≤50 mg/kg GHS acute oral toxicity category was correctly predicted
- Four (33%) of 12 substances in the 50 < LD_{50} ≤300 mg/kg GHS acute oral toxicity category were correctly predicted; however, since 10 other substances were also predicted for this category, the predictivity was 29% (4/14)
- Twelve (75%) of 16 substances in the 300 < LD_{50} ≤2000 mg/kg GHS acute oral toxicity category were predicted correctly. Since a total of 40 substances were predicted for this category, the predictivity was 30% (12/40)
- Four (40%) of 10 substances in the 2000 < LD_{50} ≤5000 mg/kg GHS acute oral toxicity category were correctly predicted; however, since a total of 11 substances were predicted for this category, the predictivity was 36% (4/11).
- Zero (0%) of 12 substances with LD_{50} >5000 mg/kg were correctly predicted

The overall accuracy of the NHK NRU test method with the RC rat-only weight regression was 31% (21/68). Toxicity was overpredicted for 37% (22) and underpredicted for 32% (25) of the 68 substances. For this analysis, in terms of each GHS acute oral toxicity classification category:

- Zero (0%) of six substances with LD_{50} <5 mg/kg were correctly predicted
- One (9%) of 11 substances in the 5 < LD_{50} ≤ 50 mg/kg GHS acute oral toxicity category was correctly predicted
- Five (42%) of 12 substances in the 50 < LD_{50} ≤300 mg/kg GHS acute oral toxicity category were correctly predicted; however, since six other substances were also predicted for this category, the predictivity was 33% (3/9)
- Thirteen (81%) of 16 substances in the 300 < LD_{50} ≤2000 mg/kg GHS acute oral toxicity category were predicted correctly; however, since 29 other

substances were also predicted for this category, the predictivity was 31% (13/42)

- One (10%) of 10 substances in the 2000 < LD_{50} ≤5000 mg/kg GHS acute oral toxicity category was correctly predicted
- One (8%) of 13 substances with LD_{50} >5000 mg/kg was correctly predicted

2.4 Test Method Reliability (Inter- And Intra-Laboratory Reproducibility)

Reproducibility is the consistency of individual test results obtained within a single laboratory (intralaboratory reproducibility) or among different laboratories (interlaboratory reproducibility) using the same protocol and test samples. Reproducibility was evaluated using the results from the reference substances that yielded IC_{50} values from all three validation study laboratories (i.e., 64 and 68 reference substances for the 3T3 and the NHK NRU test methods, respectively). Intra- and inter-laboratory reproducibility of the 3T3 and NHK NRU IC_{50} data were assessed using analysis of variance (ANOVA), coefficient of variation (CV) analysis, comparison of the laboratory-specific IC_{50}-LD_{50} regressions to one another, and comparison of maximum:minimum mean laboratory IC_{50} values. As indicated below, reproducibility was generally better for the NHK NRU test method.

Although ANOVA results for the PC, sodium lauryl sulfate (SLS), IC_{50} values for the 3T3 NRU test method indicated there were significant differences among laboratories ($p=0.006$) but not between study phases within laboratories ($p > 0.01$), the data show (see BRD Figure 7-5) that laboratory means and standard deviations from each testing phase overlap, which indicated that the IC_{50} was stable between testing phases. Interlaboratory CV values for SLS with the 3T3 NRU test method were relatively low and ranged from 2 to 16% for the various study phases. ANOVA results for the SLS IC_{50} for the NHK NRU test method also showed significant differences between laboratories ($p <0.001$) but also between study phases within laboratories ($p \leq 0.001$). A modified cell culturing method at FAL was likely responsible for SLS IC_{50} differences among the laboratories in phases Ia and Ib. Interlaboratory CV values were 39% and 21%, respectively, for phases Ia and Ib and 31% and 8%, respectively, for phases II and III. Very small but significantly different slopes ($p <0.05$; slope ranges from -0.00032 to 0.00020 for 3T3 and -0.0011 to -0.0004 for NHK) for linear regression analyses

of the SLS IC_{50} over time (within each laboratory) for both NRU test methods indicated that SLS IC_{50} was relatively stable over the 2.5 year duration of the study.

The assessment of reproducibility for reference substances by the comparisons of laboratory-specific IC_{50}-LD_{50} regressions indicated that the regressions were not significantly different from one another because the regressions for each laboratory were within the 95% confidence limits of the mean laboratory regressions. The similarity of the laboratories in LD_{50} predictions (via regression) for the reference substances is relevant with respect to the reproducibility analyses since the NRU methods are proposed for use with the regressions in determining starting doses for rodent acute oral toxicity tests.

ANOVA results for the reference substances showed significant laboratory differences for 23 substances for the 3T3 NRU test method, but only for six substances for the NHK NRU test method. Mean intralaboratory CV values were 26% for both methods, but the NHK NRU test method had a lower mean interlaboratory CV (28% vs 47% for 3T3). An analysis to determine the relationship, if any, between substance attributes and interlaboratory CV values indicated that physical form, solubility, and volatility had little effect on CV values. However, the magnitude of the CV seemed to be related to chemical class, GHS acute toxicity category, IC_{50}, and boiling point, although the usefulness of these relationships has not been established.

Mean interlaboratory CV values were larger for substances in the most toxic GHS categories than for substances in the other toxicity categories, especially with the 3T3 NRU test method. The mean interlaboratory CV for substances in the $LD_{50} \leq 5$ mg/kg (72%) and $5 < LD_{50} \leq 50$ mg/kg (78%) classes were larger than the mean overall interlaboratory CV (47%) with the 3T3 NRU test method. The mean interlaboratory NHK CV was 37% for substances with $LD_{50} \leq 5$ mg/kg, and 41% for substances with $5 < LD_{50} \leq 50$ mg/kg, while the mean overall interlaboratory CV was 28%. A Spearman correlation analysis showed that the IC_{50} was inversely correlated to interlaboratory CV for both the 3T3 ($p=0.015$) and NHK ($p=0.014$) test methods, and that boiling point was positively correlated to interlaboratory CV ($p=0.007$)

(i.e., higher boiling points were associated with higher CV values) for the 3T3 but not the NHK NRU test method (p=0.809).

The maximum:minimum mean laboratory IC_{50} values for the 3T3 NRU test method ranged from 1.1 to 21.6, with 33 (52%) of the 64 reference substances having values between 1.5 and 2.5. In contrast, the maximum:minimum mean laboratory IC_{50} values for the NHK NRU test method ranged from 1.0 to 107.6, with 50 (74%) of the 68 reference substances having values between 1.5 and 2.5.

2.5 Animal Welfare Considerations: Reduction, Refinement, and Replacement

Computer models were used to simulate the testing of the reference substances in two currently accepted sequential rodent acute oral toxicity test methods, the Up-and-Down Procedure (UDP; OECD 2001a; EPA 2002a) and the Acute Toxic Class (ATC) method (OECD 2001b) using either the default starting dose (175 mg/kg for the UDP, 300 mg/kg for the ATC), or the starting dose determined by the 3T3 and NHK NRU test methods. The simulations (10,000 per run for the UDP and 2000 per run for the ATC) were used to estimate, per substance, the number of animals that would be used and their associated survival rate. The modeling was performed using five different dose-mortality slopes[8] (i.e., 8.3, 4.0, 2.0, 0.8, and 0.5) because such slope information was not available for all of the reference substances used. To simplify the presentation of results, determination of animal use included the data for only two of the slopes, 2.0 and 8.3. The slope of 2.0 is the default used for the calculation of LD_{50} by the UDP method and the slope of 8.3 represents substances, such as pesticides, with higher slopes. Starting doses determined by either 3T3 or NHK cells were tested as were two rat-only regressions, one based on molar weight, the other on mg/kg (*in vivo*) and μg/mL (*in vitro*).

Computer simulation of the UDP testing showed that, for the substances with rat acute oral LD_{50} reference data tested in this validation study (67 for 3T3, 68 for NHK), the prediction of starting doses using the default staring dose of 175 mg/kg with the NRU test methods resulted in the use of fewer animals for UDP testing. An average of 0.49 animals (6.2%,

[8] The dose-mortality slope is the slope of the dose-response curve for mortality.

885 slope=8.3; NHK NRU test method) to 0.54 animals (5.8%, slope=2.0; 3T3 NRU test method)
886 would be saved with the RC rat-only millimole regression (**Table 2-3**). The RC rat-only
887 weight regression predicted mean animal savings of 0.54 animals (6.8%, slope=8.3; NHK
888 NRU test method) to 0.66 animals (7.0%, slope=2.0; 3T3 NRU test method) (**Table 2-4**). No
889 animal savings were predicted for substances with 50 $<LD_{50} \leq 300$ mg/kg; this category
890 includes the default starting dose of 175 mg/kg. The highest statistically significant animal
891 savings were predicted for substances with $2000 < LD_{50} \leq 5000$ mg/kg and $LD_{50} > 5000$
892 mg/kg for both NRU test methods. The greatest animal savings were observed for substances
893 in these categories because the limit test, which would be used for such substances, uses
894 fewer animals that the main test. When using the RC rat-only millimole regression, animal
895 savings for these categories ranged from 1.28 (11.9%) to 1.58 (20.3%) animals. Using the
896 RC rat-only weight regression produced animal savings of 1.28 (14.0%) to 1.65 animals
897 (16.7%) for the substances in these toxicity categories. Although using the 3T3 and NHK
898 NRU IC_{50} values to estimate starting doses for the simulated UDP deceased the number of
899 animals used, it did not change the number of animals that died.

Table 2-3 Animal Use[1] for the UDP[2] by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Millimole Regression[4]

		Dose-mortality Slope = 2.0			Dose-mortality Slope = 8.3		
GHS Acute Oral Toxicity Category[3]	**Number of Reference Substances**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose[6]**	**Animals Saved[7]**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose[6]**	**Animals Saved[7]**
3T3 NRU Test Method							
$LD_{50} \leq 5$ mg/kg	6	11.32 ± 0.20	10.19 ± 0.70	1.14 (10.0%)	9.70 ± 0.28	8.74 ±0.43	0.96 (9.9%)
$5 < LD_{50} \leq 50$ mg/kg	11	9.68 ± 0.23	9.74 ± 0.45	-0.07 (-0.7%)	8.46 ± 0.28	8.54 ± 0.47	-0.08 (-1.0%)
$50 < LD_{50} \leq 300$ mg/kg	12	7.76 ± 0.10	8.18 ± 0.21	-0.42 (-5.5%)	6.61 ± 0.19	6.90 ± 0.19	-0.29 (-4.3%)
$300 < LD_{50} \leq 2000$ mg/kg	16	8.53 ± 0.21	8.14 ± 0.21	0.38 (4.5%)	7.46 ± 0.24	7.15 ± 0.19	0.31* (4.1%)
$2000 < LD_{50} \leq 5000$ mg/kg	10	10.73 ± 0.10	9.46 ± 0.15	1.28* (11.9%)	9.17 ± 0.23	7.96 ± 0.31	1.21* (13.2%)
$LD_{50} > 5000$ mg/kg	12	9.87 ± 0.34	8.29 ± 0.49	1.58* (16.0%)	7.76 ± 0.59	6.18 ± 0.69	1.58* (20.3%)
Overall	67			-0.42 to 1.58			-0.29 to 1.58
NHK NRU Test Method							
$LD_{50} \leq 5$ mg/kg	6	11.21 ± 0.24	10.47 ± 0.71	0.75 (6.7%)	9.66 ± 0.27	8.95 ± 0.52	0.71 (7.3%)
$5 < LD_{50} \leq 50$ mg/kg	11	9.65 ± 0.16	9.99 ± 0. 45	-0.34 (-3.5%)	8.43 ± 0.26	8.77 ± 0.49	-0.33 (-3.9%)
$50 < LD_{50} \leq 300$ mg/kg	12	7.78 ± 0.11	8.12 ± 0.21	-0.34 (-4.4%)	6.57 ± 0.19	6.85 ± 0.19	-0.28 (-4.2%)
$300 < LD_{50} \leq 2000$ mg/kg	16	8.55 ± 0.22	8.03 ± 0.23	0.52* (6.1%)	7.49 ± 0.25	7.00 ± 0.20	0.49* (6.5%)
$2000 < LD_{50} \leq 5000$ mg/kg	10	10.75 ± 0.08	9.54 ± 0.20	1.21* (11.3%)	9.17 ± 0.23	8.06 ± 0.29	1.11* (12.1%)
$LD_{50} > 5000$ mg/kg	13	9.87 ± 0.32	8.41 ± 0.44	1.47* (14.8%)	7.66 ± 0.59	6.18 ± 0.69	1.47* (19.2%)
Overall	68			-0.34 to 1.47			-0.33 to 1.47

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity; UDP=Up-and-Down Procedure.

*Statistically significant ($p<0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference shown in parentheses.

[1]Mean numbers of animals used ± standard errors for 10,000 simulations for each substance with an upper limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method. Substances were categorized using the reference LD_{50} values in mg/kg.

[2]OECD (2001a); EPA (2002a).

[3]UN (2005).

[4]The RC rat-only millimole regression is $\log LD_{50}$ (mmol/kg) = 0.439 $\log IC_{50}$ (mM) + 0.621.

[5]Default starting dose = 175 mg/kg.

[6]The starting dose was one default dose lower than the predicted LD_{50} calculated using the IC_{50} value for each reference substance in the RC rat-only millimole regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the predicted starting dose.

Table 2-4 Animal Use[1] for the UDP[2] by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression[4]

		Dose-mortality Slope = 2.0			Dose-mortality Slope = 8.3		
GHS Acute Oral Toxicity Category[3]	**Number of Reference Substances**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose**	**Animals Saved[7]**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose**	**Animals Saved[7]**
3T3 NRU Test Method							
LD_{50} ≤5 mg/kg	6	11.29 ± 0. 20	10.38 ± 0.62	0.90 (8.0%)	9.70 ± 0.28	8.92 ± 0.37	0.78 (8.0%)
5 < LD_{50} ≤50 mg/kg	11	9.71 ± 0.22	9.58 ± 0.42	0.13 (1.3%)	8.47 ± 0.28	8.41 ± 0.44	0.06 (0.8%)
50 < LD_{50} ≤300 mg/kg	12	7.74 ± 0.10	7.99 ± 0.18	-0.25 (-3.3%)	6.58 ± 0.19	6.76 ± 0.18	-0.18 (-2.7%)
300 < LD_{50} ≤2000 mg/kg	16	8.52 ± 0.21	8.16 ± 0.19	0.35 (4.1%)	7.46 ± 0.24	7.17 ± 0.16	0.28* (3.8%)
2000 < LD_{50} ≤5000 mg/kg	10	10.78 ± 0.11	9.14 ± 0.24	1.64* (15.2%)	9.20 ± 0.24	7.61 ± 0.37	1.59* (17.3%)
LD_{50} >5000 mg/kg	12	9.87 ± 0.34	8.23 ± 0.48	1.65* (16.7%)	7.76 ± 0.59	6.14 ± 0.69	1.63* (21.0%)
Overall	67			-0.25 to 1.65			-0.18 to 1.63
NHK NRU Test Method							
LD_{50} ≤5 mg/kg	6	11.21 ± 0.24	10.49 ± 0.71	0.72 (6.4%)	9.66 ± 0.27	8.97 ± 0.52	0.69 (7.1%)
5 < LD_{50} ≤50 mg/kg	11	9.70 ± 0.18	9.78 ± 0.41	-0.07 (-0.8%)	8.45 ± 0.27	8.59 ± 0.44	-0.13 (-1.6%)
50 < LD_{50} ≤300 mg/kg	12	7.75 ± 0.11	7.99 ± 0.21	-0.24 (-3.1%)	6.58 ± 0.19	6.76 ± 0.18	-0.18 (-2.7%)
300 < LD_{50} ≤2000 mg/kg	16	8.54 ± 0.21	8.20 ± 0.22	0.34 (3.9%)	7.48 ± 0.23	7.17 ± 0.16	0.31 (4.1%)
2000 < LD_{50} ≤5000 mg/kg	10	10.77 ± 0.08	9.40 ± 0.25	1.38*(12.8%)	9.18 ± 0.23	7.90 ± 0.33	1.28* (14.0%)
LD_{50} >5000 mg/kg	13	9.88 ± 0.32	8.34 ± 0.44	1.54*(15.6%)	7.66 ± 0.56	6.12 ± 0.63	1.53* (20.0%)
Overall	68			-0.24 to 1.54			-0.18 to 1.53

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity; UDP=Up-and-Down Procedure.

*Statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test. Percent difference is shown in parentheses.

[1]Mean number of animals used ± standard errors for 10,000 simulations for each substance with a limit dose of 5000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances for the 3T3 NRU test method and 68 substances for the NHK NRU test method categorized using the reference LD_{50} values in mg/kg.

[2]OECD (2001a); EPA (2002a).

[3]UN (2005).

[4]The RC rat-only weight regression is log LD_{50} (mg/kg) = 0.372 log IC_{50} (µg/mL) + 2.024

[5]Default starting dose = 175 mg/kg.

[6]The starting dose was one default dose lower than the predicted LD_{50} calculated using the IC_{50} values for each reference substance in the RC rat-only weight regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the predicted starting dose.

Computer simulation of ATC testing showed that, for the substances tested in this validation study, the prediction of starting doses using the NRU test methods resulted in a savings of 0.51 animals (4.8%, slope=8.3 [3T3]) to 0.80 animals (7.3%, slope=2.0 [NHK]) per test when using the RC rat-only millimole regression (**Table 2-5**). The RC rat-only weight regression produced animal savings of 0.91 animals (8.6%, slope=8.3) to 1.09 animals (10.2%, slope=8.3) (**Table 2-6**). No animal savings were predicted for substances with 300 < LD_{50} ≤2000 mg/kg when reference substances were grouped by GHS acute oral toxicity category; this category includes the default starting dose of 300 mg/kg. Statistically significant mean animal savings for ATC testing were highest for substances with 5 < LD_{50} ≤50 mg/kg and for substances with LD_{50} >5000 mg/kg. Mean animal savings using the RC rat-only millimole regression for both test methods for substances with 5 < LD_{50} ≤50 mg/kg ranged from 1.15 animals (9.8%, slope=8.3) to 1.33 animals (11.4%, slope=8.3). Mean animal savings for substances with LD_{50} >5000 mg/kg ranged from 2.03 animals (17.1%, slope=2) to 2.66 animals (22.2%, slope=8.3). Using the RC rat-only weight regression, mean animal savings for both test methods for substances with 5 < LD_{50} ≤50 mg/kg ranged from 1.25 animals (10.8%, slope=2) to 1.51 animals (13.0%, slope=2.0). Mean animal savings for both test methods for substances with LD_{50} >5000 mg/kg ranged from 2.94 animals, (24.8%, slope=2.0) to 3.33 animals (27.7%; slope=8.3).

Animal savings did not correlate with the accuracy of the GHS acute oral toxicity category predictions based on the LD_{50} values calculated using the IC_{50} values in the RC rat-only regressions. The reason that animal savings is unrelated to the accuracy of prediction of GHS acute oral toxicity category based on the LD_{50} values calculated using IC_{50} values in the RC rat-only regressions is because two different standards were used for comparison in the two analyses:

- GHS acute oral toxicity category predictions were compared with the GHS categories derived from the *in vivo* reference rat oral LD_{50}
- The number of animals used (to determine animal savings) was compared with the animal use at the default starting dose of 175 mg/kg for the UDP or 300 mg/kg for the ATC

960 Despite the relatively poor GHS accuracy for the low toxicity chemicals (the toxicity of
961 almost all were overpredicted by one GHS category), animals were greatest due to the fact
962 that testing goes to the limit dose faster.
963

Table 2-5 Animal Savings[1] for the ATC[2] Method by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Millimole Regression[4]

		Dose-Mortality Slope = 2.0			Dose-Mortality Slope = 8.3		
GHS Acute Oral Toxicity Category[3]	**Number of Reference Substances**	**With Default Starting Dose[5]**	**With IC_{50}-Based Starting Dose[6]**	**Animals Saved[7]**	**With Default Starting Dose[5]**	**WithIC_{50}-Based Starting Dose[6]**	**Animals Saved[7]**
3T3 NRU Test Method							
LD_{50} ≤5 mg/kg	6	9.77 ± 0.17	7.09 ± 1.09	2.68 (27.4%)	9.08 ± 0.08	6.38 ± 1.09	2.70 (29.7%)
5 < LD_{50} ≤50 mg/kg	11	11.56 ± 0.21	10.39 ± 0.52	1.17* (10.2%)	11.75 ± 0.16	10.60 ± 0.43	1.15* (9.8%)
50 < LD_{50} ≤300 mg/kg	12	10.81 ± 0.20	10.39 ± 0.17	0.42 (3.9%)	9.42 ± 0.26	9.27 ± 0.11	0.15 (1.6%)
300 < LD_{50} ≤2000 mg/kg	16	9.75 ± 0.07	10.67 ± 0.48	-0.92* (-9.5%)	9.26 ± 0.10	10.56 ± 0.62	-1.30* (-14.0%)
2000 < LD_{50} ≤5000 mg/kg	10	11.22 ± 0.08	11.14 ± 0.08	0.08 (0.7%)	11.88 ± 0.10	11.77 ± 0.10	0.11 (0.9%)
LD_{50} >5000 mg/kg	12	11.85 ± 0.04	9.82 ± 0.78	2.03* (17.1%)	12.00 ± 0.000	9.81 ± 0.84	2.19* (18.3%)
Overall	67			-0.92 to 2.68			-1.30 to 2.70
NHK NRU Test Method							
LD_{50} ≤5 mg/kg	6	9.74 ± 0.16	6.78 ± 1.31	2.96 (30.4%)	9.09 ± 0.08	6.09 ± 1.23	2.99 (33.0%)
5 < LD_{50} ≤50 mg/kg	11	11.56 ± 0.21	10.38 ± 0.35	1.18* (10.2%)	11.76 ± 0.17	10.42 ± 0.45	1.33* (11.4%)
50 < LD_{50} ≤300 mg/kg	12	10.83 ± 0.21	10.39 ± 0.29	0.44 (4.0%)	9.44 ± 0.26	9.63 ± 0.49	-0.20 (-2.1%)
300 < LD_{50} ≤2000 mg/kg	16	9.77 ± 0.06	10.37 ± 0.49	-0.60 (-6.1%)	9.26 ± 0.10	10.11 ± 0.63	-0.85 (-9.2%)
2000 < LD_{50} ≤5000 mg/kg	10	11.22 ± 0.08	11.25 ± 0.12	-0.03 (-0.3%)	11.87 ± 0.10	11.89 ± 0.15	-0.02 (-0.2%)
LD_{50} >5000 mg/kg	13	11.86 ± 0.03	9.43 ± 0.73	2.43* (20.5%)	12.00 ± 0.000	9.34 ± 0.80	2.66* (22.2%)
Overall	68			-0.60 to 2.96			-0.85 to 2.99

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity.

*Statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.

[1]Mean number of animals used ± standard errors for 2000 simulations for each substance with an upper limit dose of 2000 mg/kg. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method categorized using the reference LD_{50} values in mg/kg from BRD **Table 4-2**. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers.

[2]OECD (2001d).

[3]GHS for acute oral toxicity (UN 2005).

[4]The RC rat-only millimole regression is log LD_{50} (mmol/kg) = 0.439 log IC_{50} (mM) + 0.621.

[5]Default starting dose =300 mg/kg.

[6] The starting dose was the next fixed dose lower than the predicted LD_{50} using the IC_{50} for each reference substance in the RC rat-only millimole regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose.

Table 2-6 Animal Savings[1] for the ATC[2] Method by GHS Acute Oral Toxicity Category[3] Using Starting Doses Based on the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression[4]

GHS Acute Oral Toxicity Category[3]	Number of Reference Substances	Dose-Mortality Slope = 2.0			Dose-Mortality Slope = 8.3		
		With Default Starting Dose[5]	With IC_{50}-Based Starting Dose[6]	Animals Saved[7]	With Default Starting Dose[5]	With IC_{50}-Based Starting Dose[6]	Animals Saved[7]
3T3 NRU Test Method							
LD_{50} ≤5 mg/kg	6	9.77 ± 0.17	7.56 ± 1.03	2.21 (22.6%)	9.08 ± 0.08	6.85 ± 0.99	2.24 (24.6%)
5 < LD_{50} ≤50 mg/kg	11	11.56 ± 0.21	10.06 ± 0.38	1.51* (13.0%)	11.75 ± 0.16	10.27 ± 0.33	1.48* (12.6%)
50 < LD_{50} ≤300 mg/kg	12	10.81 ± 0.20	10.35 ± 0.18	0.47* (4.3%)	9.42 ± 0.26	9.20 ± 0.10	0.22 (2.4%)
300 < LD_{50} ≤2000 mg/kg	16	9.75 ± 0.07	10.67 ± 0.50	-0.93* (-9.5%)	9.26 ± 0.10	10.65 ± 0.66	-1.39 (-15.0%)
2000 < LD_{50} ≤5000 mg/kg	10	11.22 ± 0.08	9.80 ± 0.51	1.43* (12.7%)	11.88 ± 0.10	9.44 ± 0.88	2.43 (20.5%)
LD_{50} >5000 mg/kg	12	11.85 ± 0.04	8.83 ± 0.83	3.02* (25.5%)	12.00 ± 0.00	8.67 ± 0.91	3.33* (27.7%)
Overall	67			-0.93 to 3.02			-1.39 to 3.33
NHK NRU Test Method							
LD_{50} ≤5 mg/kg	6	9.74 ± 0.16	6.87 ± 1.28	2.87 (29.4%)	9.09 ± 0.08	6.18 ± 1.20	2.91 (32.0%)
5 < LD_{50} ≤50 mg/kg	11	11.56 ± 0.21	10.31 ± 0.19	1.25* (10.8%)	11.76 ± 0.17	10.40 ± 0.33	1.36* (11.5%)
50 < LD_{50} ≤300 mg/kg	12	10.83 ± 0.21	10.41 ± 0.28	0.42 (3.8%)	9.44 ± 0.26	9.63 ± 0.49	-0.20 (-2.1%)
300 < LD_{50} ≤2000 mg/kg	16	9.77 ± 0.62	10.46 ± 0.50	-0.69 (-7.1%)	9.26 ± 0.10	10.23 ± 0.65	-0.97 (-10.4%)
2000 < LD_{50} ≤5000 mg/kg	10	11.22 ± 0.09	10.69 ± 0.37	0.53 (4.7%)	11.87 ± 0.10	11.03 ± 0.60	0.84 (7.1%)
LD_{50} >5000 mg/kg	13	11.86 ± 0.03	8.91 ± 0.78	2.94* (24.8%)	12.00 ± 0.00	8.75 ± 0.85	3.25* (27.1%)
Overall	68			-0.69 to 2.94			-0.97 to 3.25

Abbreviations: 3T3=BALB/c 3T3 fibroblasts; ATC=Acute Toxic Class method; GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NHK=Normal human keratinocytes; RC=Registry of Cytotoxicity.

*Statistically significant ($p < 0.05$) by a one-sided Wilcoxon signed rank test. Percentage difference is shown in parentheses.

[1]Mean number of animals used ± standard errors for 2000 simulations for each substance with an upper limit dose of 2000 mg/kg. Although the simulations used whole animals, averaging the results over a large number of simulations produced fractional numbers. Results are provided for 67 substances in the 3T3 NRU test method and 68 substances in the NHK NRU test method categorized using the reference LD_{50} values in mg/kg.

[2]OECD (2001d).

[3]GHS for acute oral toxicity (UN 2005).

[4]$\log LD_{50}$ (mg/kg) = 0.372 $\log IC_{50}$ (μg/mL) + 2.024

[5]Default starting dose = 300 mg/kg.

[6] The starting dose was one fixed dose lower than the predicted LD_{50} calculated using the IC_{50} for each reference substance in the RC rat-only weight regression. The IC_{50} value for each reference substance was randomly selected from the distribution of values obtained during the testing with each method.

[7]Difference between mean animal use with the default starting dose and mean animal use with the IC_{50}-based starting dose

2.6 ICCVAM Recommendations for Test Method Uses

ICCVAM's recommendations for use of these test methods is as follows:

1. The 3T3 and NHK NRU test methods are not sufficiently accurate to predict acute oral toxicity for the purpose of regulatory hazard classification (see **Section 2.3** above and **Section 6** of the *In Vitro* Acute Toxicity Test Methods BRD).
2. For the purposes of acute oral toxicity testing, the 3T3 and NHK NRU test methods may be used in a weight-of-evidence approach to determine the starting dose for the current acute oral toxicity protocols (i.e., the UDP, the ATC).
3. Consistent with the U.S. Government Principles on the Use of Animals in Research, Testing, and Education[9], and the U.S. Public Health Service Policy on Humane Care and Use of Laboratory Animals (PHS 2002), *in vitro* basal cytotoxicity test methods as part of a weight-of-evidence approach to estimate the starting dose for acute oral *in vivo* toxicity test methods should be considered and used where appropriate before testing is conducted using animals. For some types of substances, this approach will reduce the number of animals needed. In some testing situations, the approach may also reduce the numbers of animals that die or need to be humanely killed**.**
4. The starting doses for substances with certain toxic mechanisms that are not expected to be active in 3T3 or NHK cells (e.g., those that are neurotoxic or cardiotoxic) will likely be underpredicted by these *in vitro* basal cytotoxicity test methods. Therefore, the results from basal cytotoxicity testing with such substances may not be appropriate for estimating starting doses.
5. The regression formula used to determine starting doses for test substances with known molecular weights and high purity should be the revised RC millimole regression line, based on substances with rat LD_{50} data, with IC_{50} values in mmol/L and LD_{50} values in mmol/kg. The regression formula used

[9] IRAC (Interagency Research Animal Committee). 1985. U.S. Government Principles for Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training. Federal Register, 1985, May 20, Vol. 50, No.97.

to determine starting doses for mixtures, test substances with low or unknown purity, or test substances with unknown molecular weights should be the revised RC regression line, based on substances with rat LD_{50} data, with IC_{50} values in μg/mL and LD_{50} values in mg/kg.

6. The performance of other *in vitro* basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD_{50} value, respectively) should be demonstrated to meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods.
7. Compared to the NHK NRU test method, the 3T3 NRU test method appears to be less labor intensive and less expensive to conduct; therefore, the 3T3 NRU test method is recommended for general use. Although the 3T3 NRU test method was less reproducible than the NHK NRU test method, it produced slightly higher animal savings and accuracy for prediction of GHS acute oral toxicity category using the IC_{50} and the revised RC regressions evaluated for the prediction of LD_{50}.

2.7 ICCVAM Future Study Recommendations

ICCVAM recommends the following future studies in order to advance the use of *in vitro* methods for assessing acute oral toxicity for regulatory hazard classification purposes:

1. Additional data should be collected using the 3T3 NRU basal cytotoxicity test method to evaluate its usefulness for predicting the rodent acute oral toxicity of chemical mixtures.
2. To supplement the high quality validation database started by this study, additional high quality comparative *in vitro* basal cytotoxicity data should be collected when rat acute oral toxicity testing is conducted. However, *in vivo* testing should not be conducted solely to collect data to assess the usefulness of the NRU test method. Periodic evaluations of the expanded database should be conducted to further characterize the usefulness and limitations of using *in vitro* cytotoxicity data as part of a weight-of-evidence approach to estimate starting doses.

3. Additional efforts should be conducted to identify *in vitro* tests and other methods necessary to achieve accurate acute oral hazard classification; studies should be conducted to investigate the potential use of *in vitro* cell-based test methods that incorporate mechanisms of action and evaluations of ADME (absorption, distribution, metabolism, excretion) to provide improved estimates of acute toxicity hazard categories. Methods should be developed to extrapolate from *in vitro* toxic concentrations to equivalent doses *in vivo*.
4. The *in vivo* database of reference substances used in this validation study should be used to evaluate the utility of other non-animal approaches to estimate starting doses for acute oral toxicity tests (e.g., widely available software that uses quantitative structure-activity relationships [QSAR]).
5. Standardized procedures to collect *in vivo* measurements and observations pertinent to an understanding of the mechanisms of lethality should be included in future rat acute oral toxicity studies. Such information will likely be necessary to support the further development of predictive mechanism-based *in vitro* methods.
6. An expanded list of reference substances with rat acute oral LD_{50} values substantiated by high quality *in vivo* data (including data currently held by industry) should be developed for use in future *in vitro* test method development and validation studies.

3.0 ICCVAM RECOMMENDED PERFORMANCE STANDARDS

The purpose of performance standards is to communicate the basis by which validated new proprietary (e.g., copyrighted, trademarked, registered) and nonproprietary test methods have been determined to have sufficient accuracy and reliability for specific testing purposes. Performance standards can then be used to evaluate the accuracy and reliability of other test methods that are based on similar scientific principles and that measure or predict the same biological or toxic effect. The three elements of performance standards are essential test method components (see **Section 3.1**), a minimum list of reference substances for assessing the accuracy and reliability of the proposed test method (see **Section 3.2**), and the accuracy and reliability values that should be achieved by the proposed test method using the minimum list of reference substances (see **Section 3.3**).

The 3T3 and NHK NRU test methods are not sufficiently accurate to predict the acute oral toxicity of substances for the purposes of regulatory hazard classification and labeling. However, these test methods may be used in a weight-of-evidence approach to determine the starting dose for the UDP (OECD 2001a; EPA 2002a) and the ATC (OECD 2001b) rodent acute oral toxicity test methods. The performance of other *in vitro* basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD_{50}, respectively) should meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods.

The extent to which proposed *in vitro* basal cytotoxicity test methods should demonstrate comparable performance to these two *in vitro* NRU cytotoxicity test methods should be considered on a case-by-case basis.

3.1 Essential Test Method Components for *In Vitro* Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity Tests

These consist of essential structural, functional, and procedural elements of a validated test method that should be included in the protocol of a proposed, mechanistically and functionally similar test method. Essential test method components include unique characteristics of the test method, critical procedural details, and quality control measures.

Adherence to essential test method components will help to assure that a proposed test method is structurally and functionally similar to the corresponding validated test method.

The basic steps of an *in vitro* basal cytotoxicity assay are as follows:

- The test substance is dissolved in an appropriate solvent and applied as a solution to cells that, under control conditions, would be expected to be growing exponentially throughout the exposure period.
- The test substance is incubated with the cells for a specified period of time.
- The test substance is removed and an endpoint indicative of cell viability or cytotoxicity is measured.
- The IC_{50} value is calculated (i.e., the concentration at which cell viability or growth is inhibited by 50% compared to control values).

Many different *in vitro* basal cytotoxicity methods might be used to estimate rat acute oral LD_{50} values and, thus, to predict the starting dose for a rodent acute oral lethality assay. *In vitro* basal cytotoxicity data determined using various primary cells and permanent non-differentiated finite or transformed cell lines, generally exhibit the same concentration-response cytotoxicity relationship when exposed to the same xenobiotic, regardless of the toxic endpoints investigated. The following endpoints are sufficiently characteristic of basal cytotoxicity (Spielmann et al. 1999; Halle 1998, 2003):

- Inhibition of cell proliferation: cell number, cell protein, deoxyribonucleic acid (DNA) content, DNA synthesis, colony formation
- Cell viability - metabolic markers: metabolic inhibition test, mitochondrial reduction of tetrazolium salts into soluble dye
- Decreased cell viability - membrane markers: NRU into cell lysosomes, Trypan Blue exclusion, cell attachment/cell detachment for monolayer cultures
- Differentiation markers: functional or morphological differentiation within cell clusters, intracellular morphology

Markers of the release of intracellular components, such as the enzyme lactate dehydrogenase (i.e., LDH release test) or of dye introduced into the cells previous to chemical exposure as occurs, for example, in the fluorescein leakage (FL) test or the Neutral Red Release (NRR) test, are not considered to be characteristic for basal cytotoxicity because they specifically detect damage of the outer cell membrane and generally are associated with short-term chemical exposure (ICCVAM 2001). A chemical that specifically damages only cell membranes, however, will be detected correctly in one of the tests for basal cytotoxicity listed above.

Investigators using an *in vitro* basal cytotoxicity system for prediction of the *in vivo* starting dose for acute oral toxicity studies must be able to demonstrate that the assay is valid for its intended use. This includes demonstrating that any modification to the existing validated reference test method does not adversely affect its performance characteristics. *In vitro* systems may be used to test solids, liquids, and emulsions of any chemical or product class. The liquids can be aqueous or nonaqueous; solids can be soluble or insoluble in water. The samples may be pure chemicals, dilutions, formulations, or waste. Test substances must be soluble in cell culture medium, dimethyl sulfoxide (DMSO), or ethanol (ETOH). The test method endpoint (i.e., percent of control values) is used to generate an IC_{50} value in mM (if the substance's molecular weight is known, and, if not, in μg/mL) and the IC_{50} value is used in the regressions developed to estimate the LD_{50} value in mmol/kg (or mg/kg).

The following is a description of the essential test method components for *in vitro* basal cytotoxicity assays to predict starting doses for acute oral toxicity/lethality tests.

3.1.1 *In Vitro* Cell Culture Conditions

- A mammalian cell line (or primary cells) is used that divides rapidly with doubling times of less than 30 hours under standard culture conditions, preferably with calf serum [CS], newborn calf serum [NCS]), or serum-free medium (ICCVAM 2001).
- Cells are allowed to propagate in sterile tissue culture vessels (e.g., flasks) and then are subcultured to other sterile tissue culture vessels (e.g., 96 well-plates)

for use in testing. Initial cell seeding should be done at a density that allows for exponential growth throughout the exposure period.

- Appropriate cell culture growth conditions are maintained throughout the testing period (e.g., 37°C ± 1°C, 90% ± 10% humidity, 5.0% ± 1% CO_2/air). The cell cultures should be free of contamination with bacteria, mycoplasma, or fungi.
- Cell culture media should be prequalified by the testing laboratory via a standardized protocol before initiating the test to guarantee that the media provide cells with appropriate nutrients to meet the growth criteria required for the test method.

3.1.2 Application of the Test Substances

Test Substance Preparation

- Test substance solutions should be prepared in cell culture medium within an hour before application to the cell cultures (unless the stability of the test substance in the solvent used requires shorter times or allows longer times).
- Standard protocol methods for solubility procedures can include mixing the test substance by vortexing, sonication, warming, and stirring. Test substances should be fully solubilized (i.e., no visual observation of test substance in the dosing solution) before application.
- An inherent limitation to *in vitro* cytotoxicity is the testing of volatile substances since the material may evaporate before application to the cells or may not remain in the test vessel when incubated. If volatility is predicted or identified for a test substance (e.g., by detection of cross-contamination of the high concentrations of test substance in culture with lower concentrations or controls in the test vessel), measures can be employed to test moderately volatile substances (e.g., cover the test plate with a CO_2 permeable plastic film cover/sealer).

Cytotoxicity Test

- Each cytotoxicity test should contain a range of test substance concentrations such that the IC_{50} value can be determined with at least one cytotoxic point between 0 – 50% viability and at least one cytotoxic point between 50 – 100% viability.
- A minimum of three adequate data points should be collected for each test substance concentration. (Note: The NICEATM/ECVAM validation study required the testing of six replicates for each test substance concentration with at least four successful replicates.)
- Blanks (i.e., culture vessels without cells) should be available for assessing background interference when measuring the endpoint.
- Cell monolayers in tissue culture vessels should be adequately covered (e.g., a minimum of 100 μL of test substance solution per well in a 96-well test plate).
- The substance exposure period should be at least the duration of one cell cycle (i.e., approximately 24 to 72 hours) (Riddell et al. 1986). [Note: The NICEATM/ECVAM validation study required an exposure period of 48 hours for 3T3 and NHK cells; the cell cycle duration (i.e., doubling time) for these cells ranged from 17 to 19 (3T3) and 10 to 22 (NHK) hours in log phase.]
- At the end of the exposure period, most endpoints require washing the test substance from the cells with an appropriate buffering solution (e.g., Dulbecco's Phosphate Buffered Saline [DPBS]) before applying the endpoint material (e.g., neutral red dye). Washing cells to remove the test substance is the default recommendation unless it is known that washing would interfere with measurement of the endpoint.

3.1.3 Control Substances

Vehicle Controls (VC): The VCs provide the reference for 100% cell growth in the test vessel and, thus, the vehicle (or solvent) must be compatible with the cell culture system (i.e., not cause cytotoxicity or reduce cell growth through other mechanisms) and should not alter the properties of the test substance. The VCs should contain the solvent at the concentration applied to the cells. For example, DMSO and ETOH at a final concentration ≤ 0.5% [v/v] were demonstrated to be compatible with cell growth for 3T3 and NHK cells in the

NICEATM/ECVAM validation study. If the compatibility of the solvent with the cell culture system is unknown, cultures with and without the solvent should be included in each experiment.

Positive Controls (PC): The purpose of a PC substance is to demonstrate that the cell culture system is responding with adequate sensitivity to a cytotoxic agent for which the magnitude of the cytotoxic response is well characterized. The PC substance should be tested concurrently with (and independent of) the test substance. The PC should be well characterized for its cytotoxicity potential and each test should generate a response that is comparable to the historic IC_{50} range generated by the laboratory. A laboratory should perform a minimum of 10 cytotoxicity tests using the PC over a number of days to develop a minimum historical database of IC_{50} data. Typically, for biologically based test methods, suggested acceptable ranges for the PC response are within two to three standard deviations of the historical mean response, but developers of proprietary test methods may establish tighter ranges. Sodium lauryl sulfate (SLS) is an effective PC substance for use in *in vitro* basal cytotoxicity test methods. [Note: The NICEATM/ECVAM validation study used SLS as the PC and required 2.5 standard deviations of the historical mean response as the acceptable range.]

Benchmark Controls: Benchmark controls may be useful to demonstrate that the test method is functioning properly for detecting the cytotoxic potential of substances of a specific chemical class or a specific range of responses, or for evaluating the relative cytotoxic potential of a cytotoxic test substance. Appropriate benchmark controls should have the following properties:

- Consistent and reliable source(s) for the substance
- Structural and functional similarity to the class of the substance being tested
- Known physical/chemical characteristics
- Supporting data on known effects in animal models
- Known potency in the range of response (including moderate response)

3.1.4 Viability Measurements

- Only standardized, quantitative methods should be used to measure cell viability. The protocol should be compatible with laboratory apparatus such as spectrophotometers that allow a quick and precise measurement of the endpoint.
- Non-specific dye binding must not interfere with the viability measurement. A measurement endpoint that is well established and that has good interlaboratory reproducibility should be used (ICCVAM 2001).
- A detailed concentration-response experiment should be conducted using a progression factor that yields graded effects between no effect and total cytotoxicity. Any desired toxicity measure can be derived from a well-designed concentration-response experiment.
- Preference should be given to endpoints that determine either cell proliferation or cell viability (e.g., NRU, MTT [3-(4,5,dimethylthiazol-2yl)2,5-diphenyl tetrazolium bromide], XTT [Sodium 3,3,-[(phenylamino)carbonyl]-3,4-tetrazolium-bis(4-methoxy-6-nitro)benzenesulfonic acid hydrate]) (ICCVAM 2001).
- Simple endpoints such as total protein content are not recommended, as they may under-predict the toxicity of certain test substances by including protein from dead cells.
- A lack of information and a low level of accuracy characterize experiments that seek only to identify the highest tolerated dose or the lowest cytotoxic dose.

Colorimetric endpoints (e.g., NRU) should have the optical density (OD) spectrascopically-measured at the appropriate wavelength (e.g., 540 nm ± 10 nm for NRU) and OD values for blanks should be subtracted from the vehicle control and test substance ODs.

3.1.5 Interpretation of Results

IC_{50} Determination: The endpoint values obtained at each concentration of the test substance can be used to calculate the percentage of cell viability or growth relative to the negative (vehicle) control, which is arbitrarily set at 100%. The cell viability criteria used to determine an IC_{50} value must be clearly defined and documented, and be shown to be appropriate. In general, such criteria are established during test optimization, tested during a prevalidation phase, and confirmed in a validation study.

Regression Formula: The recommended regression formulas to predict LD_{50} values from IC_{50} values are

- The RC rat-only millimole regression, log LD_{50} mmol/kg = 0.439 log IC_{50} mM + 0.621, for substances with known molecular weight
- The RC rat-only weight regression, log LD_{50} mg/kg = 0.372 log IC_{50} μg/mL + 2.024, for mixtures and substances with no known molecular weight:

3.1.6 Test Report

The test report should include the following information, if relevant to the conduct of the study:

Test Substances and Control Substances

- Chemical name(s) such as Chemical Abstracts Service Registry Number (CASRN) and molecular weight (if known), followed by other names, if known
- Formulation (if available) of the test substance if the material is a mixture
- Purity and composition of the substance or preparation (in percentage[s] by weight)
- Physicochemical properties such as physical state, volatility, pH, stability, chemical class, water solubility relevant to the conduct of the study
- Treatment of the test/control substances prior to testing, if applicable (e.g., vortexing, sonication, warming; solvent used)
- Stability, if known

Justification of the In Vitro Test Method and Protocol Used

Test Method Integrity

- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time
- If the test method employs proprietary components, documentation on the procedure used to ensure their integrity from "lot-to-lot" and over time
- The procedures that the user may employ to verify the integrity of the proprietary components

Criteria for an Acceptable Test

- Acceptable concurrent positive control ranges based on historical data
- Acceptable negative and solvent/vehicle control data

Test Conditions

- Cell system used
- Calibration information for measuring device used for measuring cell viability (e.g., spectrophotometer)
- Details of test procedure used
- Test doses used
- Description of any modifications of the test procedure
- Reference to historical data of the model
- Description of evaluation criteria used

Results

- Tabulation of data from individual test samples (e.g., OD values and calculated percentage cell viability data for the test substance and the positive, negative, and benchmark controls, reported in tabular form, including data from replicate repeat experiments as appropriate, and means and ± the standard deviation for each trial)
- Calculated IC_{50} value
- Calculated starting dose (i.e., LD_{50} value) using IC_{50} value in regression formula
- Regression formula (prediction model) used

Description of Other Effects Observed

Discussion of the Results

Conclusion

3.2 Reference Substances for *In Vitro* Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity Tests

Reference substances are used to assess the accuracy and reliability of a proposed, mechanistically and functionally similar test method and are a representative subset of those used to demonstrate the reliability and the accuracy of the validated test method. These substances:

- Are representative of the range of responses that the validated test method is capable of measuring or predicting
- Have produced consistent results in the validated test method
- Will reflect the accuracy of the validated test method
- Have well-defined chemical structures
- Are readily available
- Are not associated with excessive hazard or prohibitive disposal costs

The subset of 30 reference substances in **Table 3-1** was chosen from the 72 reference substances used in the NICEATM/ECVAM validation study. Reference substances that exhibited solubility difficulties or were volatile in culture during this study are included as a secondary subset and are recommended for investigational purposes only.

The substances in this list represent the following types of chemical classes: acyclic hydrocarbons; alcohols; amides; amines; arsenical compounds; boron compounds; cadmium compounds; carboxylic acids; chlorine compounds cyclic hydrocarbons; fluorine compounds, heterocyclics; mercury compounds; nitro compounds; organometallics; phenols, organophosphorous compounds; polycyclics; potassium compounds; sodium compounds; and sulfur compounds, and ureas.

1370 **Table 3-1 Recommended Reference Substances for Evaluation of *In Vitro* Basal Cytotoxicity Methods for Predicting the**
1371 **Starting Dose for Rodent Acute Oral Toxicity Tests**

1372

Reference Substance	CASRN[1]	Rodent Oral LD_{50}[2]		3T3 IC_{50}[3]		NHK IC_{50}[3]	
		mg/kg	mmole/kg	µg/mL	mM	µg/mL	mM
$LD_{50} \leq 5$ mg/kg							
Mercury II chloride	7487-94-7	1	0.0037	4.122	0.0152	5.796	0.0213
Triethylenemelamine	51-18-3	1	0.0049	0.2722	0.0013	1.853	0.0091
Cycloheximide	66-81-9	2	0.0071	0.1874	0.0007	0.0734	0.0003
Busulfan	55-98-1	2	0.0081	77.68	0.3154	260.1	1.056
Phenylthiourea	103-85-5	3	0.0197	78.98	0.5189	336.3	2.210
$5 < LD_{50} \leq 50$ mg/kg							
Dichlorvos	62-73-7	17	0.0769	17.74	0.0803	10.69	0.0484
Digoxin	20830-75-5	18	0.0230	445.5	0.5705	0.0010	0.000001
Sodium arsenite	7784-46-5	41	0.3156	0.7587	0.0058	0.4766	0.0037
Triphenyltin hydroxide	76-87-9	44	0.1199	0.0172	0.00005	0.0101	0.00003
Sodium dichromate dihydrate	7789-12-0	50	0.1908	0.5867	0.0020	0.7117	0.0024
$50 < LD_{50} \leq 300$ mg/kg							
Hexachlorophene	70-30-4	61	0.1499	4.195	0.0103	0.0289	0.00007
Cadmium II chloride	10108-64-2	88	0.4801	0.5177	0.00280	1.797	0.0098
Sodium oxalate	62-76-0	155	1.160	37.14	0.2772	339.4	2.533
Sodium fluoride	7681-49-4	180	4.290	78.02	1.858	48.90	1.164
Diquat dibromide monohydrate	6385-62-2	231	0.6714	8.040	0.0222	4.333	0.0120
$300 < LD_{50} \leq 2000$ mg/kg							
Amitriptyline HCl	549-18-8	361	1.150	7.054	0.0225	8.959	0.0286
Propranolol HCl	3506-09-0	470	1.589	14.11	0.0477	36.20	0.1224
Atropine sulfate monohydrate	5908-99-6	639	0.9204	76.03	0.1094	81.83	0.1178
Acetylsalicylic acid	50-78-2	1000	5.549	676.4	3.754	605.5	3.360

Reference Substance	CASRN[1]	Rodent Oral LD_{50}[2]		3T3 IC_{50}[3]		NHK IC_{50}[3]	
		mg/kg	mmole/kg	µg/mL	mM	µg/mL	mM
Carbamazepine	298-46-4	1957	8.282	103.2	0.4367	83.24	0.3523
2000 < LD_{50} ≤ 5000 mg/kg							
Acetaminophen	103-90-2	2404	15.90	47.66	0.3152	518.0	3.426
Potassium chloride	7447-40-7	2602	34.90	3555	47.68	2237	30.01
Chloramphenicol	56-75-7	3393	10.50	130.2	0.4029	345.0	1.068
Lactic acid	50-21-5	3730	41.41	3044	33.79	1304	14.48
Trichloroacetic acid	76-03-9	4999	30.59	901.8	5.519	413.3	2.529
LD_{50} > 5000 mg/kg							
Ethylene glycol	107-21-1	8567	138.0	24435	393.6	42097	678.1
Gibberellic acid	77-06-5	6305	18.20	7810	22.55	2856	8.246
Sodium hypochlorite	7681-52-9	10328[4]	138.7[4]	1040	13.97	1502	20.18
Dibutyl phthalate	84-74-2	11998	43.11	43.37	0.1558	28.69	0.1031
Glycerol	56-81-5	12691	137.8	24345	264.4	24730	268.5
Secondary Subset							
Precipitating Substances[5]							
LD_{50} ≤ 5 mg/kg							
Arsenic trioxide	1327-53-3	20	0.1000	2.072	0.0105	6.840	0.0346
Parathion	56-38-2	2	0.0069	37.42	0.1285	30.26	0.1039
Volatile Substances[6]							
300 < LD_{50} ≤ 2000 mg/kg							
Phenol	108-95-2	414	4.400	66.32	0.7047	75.03	0.7972
LD_{50} > 5000 mg/kg							
Ethanol	64-17-5	14008	304.15	6523	141.6	10018	217.5
2-Propanol	67-63-0	5843	97.21	3489	58.04	5364	89.24

[1]Chemical Abstracts Service Registry Number
[2]The calculated value of the oral dose that produces lethality in 50% of test animals (rats and mice). Values used in the RC (Halle 1998, 2003) unless otherwise noted.
[3]Reference substance concentration (geometric mean of laboratory means) producing 50% inhibition of the endpoint measured (i.e., cell viability).
[4]LD_{50} values were calculated as the geometric mean of values obtained in the literature (see BRD Section 4).
[5]Reference substances expected to precipitate at cytotoxic concentrations.
[6]Reference substances expected to contaminate neighboring wells at high concentrations.

3.3 Accuracy and Reliability Standards

The third element of the performance standards is the determination of accuracy (also known as relevance) and reliability values.

3.3.1 Accuracy and Reliability for the NRU Test Methods

To demonstrate technical proficiency with the validated 3T3 or NHK NRU test method, ICCVAM recommends that the user evaluate his/her ability to calculate IC_{50} values for a minimum of two unclassified substances and two from each from the five GHS hazard categories (i.e., at least 12 of the 30 reference substances) listed in **Table 3-1.** The resulting IC_{50} values should be within 2.5 standard deviations of the IC_{50} values reported in the table.[10] A linear regression calculated using the LD_{50} values provided in **Table 3-1** and the resulting IC_{50} values should not differ from a linear regression calculated using the LD_{50} and the IC_{50} values provided in **Table 3-1**. Also, the intralaboratory CV values for the IC_{50} of the reference substances selected should not exceed 129% for the NHK NRU test method or 98% for the 3T3 NRU test method and the mean CV should not exceed 27% for either test method.

3.3.2 Accuracy and Reliability for Me-Too Assays

A proposed test method that is functionally and mechanistically similar to the 3T3 NRU test method should use the selected reference substances to assess accuracy and reliability. The ICCVAM Recommendations (see **Section 2.6**) propose the general use of the 3T3 NRU test method because it appears to be less labor intensive and less expensive to conduct compared to the NHK NRU test method. Thus, the accuracy and reliability standards presented below focus on the 3T3 NRU test method.

Before using a candidate *in vitro* basal cytotoxicity test to predict starting doses, the correlation between the *in vitro* and the *in vivo* test methods must be established quantitatively by using the new test method to test 12 of the 30 reference substances. After testing, the IC_{50} data are used to calculate a linear regression formula (least square method) for the selected reference substances using the corresponding LD_{50} values provided in **Table**

[10] Replicate IC_{50} values must be determined for each reference substance in order to calculate the standard deviation.

3-1. The resulting regression is compared against a regression using the 3T3 NRU IC_{50} and the LD_{50} values provided in this table. If the regressions are not statistically significantly different based on a comparison of slope and intercept (at $p < 0.05$), then the test is considered suitable to generate IC_{50} data to use with the recommended regression formula for estimating starting doses for acute oral toxicity/lethality tests.

The overall accuracy of the 3T3 NRU test method for correctly predicting GHS acute oral toxicity classification category of the 30 reference substances using the RC rat-only millimole regression was 33%. *In vivo* toxicity was overpredicted for 33% and underpredicted for 34%. Seventy-seven percent of the reference substances were classified within the correct category, or within one category above or below the correct category (see **Table 3-2**). For this analysis, in terms of each GHS acute oral toxicity classification category:

- Zero (0%) of 5 substances with $LD_{50} < 5$ mg/kg was correctly predicted
- One (20%) of 5 substances in the $5 < LD_{50} \leq 50$ mg/kg category was correctly predicted
- Four (80%) of 5 substances in the $50 < LD_{50} \leq 300$ mg/kg category were correctly predicted
- Four (80%) of 5 substances in the $300 < LD_{50} \leq 2000$ mg/kg category were correctly predicted; however, this toxicity category was also predicted for 11 other substances that did not match this category *in vivo*. Thus, the predictivity for this category was 27%.
- Zero (0%) of the 5 substances in the $2000 < LD_{50} \leq 5000$ mg/kg category were correctly predicted
- One (20%) of the 5 substances with $LD_{50} > 5000$ mg/kg were correctly predicted. The predictivity for this category was 27%.

Table 3-2 Prediction of GHS Acute Oral Toxicity Category by the 3T3 NRU Test Method Using the Recommended Reference Substances and the RC Rat-Only Millimole Regression[1]

Reference Rodent Oral LD_{50}[2] (mg/kg)	NRU-Predicted GHS Category (mg/kg)						Total	Accuracy	Toxicity Over-predicted	Toxicity Under-predicted
	LD_{50} <5	5< LD_{50} ≤50	50 < LD_{50} ≤300	300 < LD_{50} ≤2000	2000 < LD_{50} ≤5000	LD_{50} >5000				
LD_{50} < 5	0	2	1	2	0	0	5	0%	0%	100%
5< LD_{50} ≤50	0	1	2	1	1	0	5	20%	0%	80%
50 < LD_{50} ≤300	0	0	4	1	0	0	5	80%	0%	58%
300 < LD_{50} ≤2000	0	0	1	4	0	0	5	80%	20%	0%
2000 < LD_{50} ≤5000	0	0	0	5	0	0	5	0%	100%	0%
LD_{50} >5000	0	0	0	2	2	1	5	20%	80%	0%
Total	0	3	8	15	3	1	30	33%	33%	34%
Predictivity	0%	33%	50%	27%	0%	100%				
Category Overpredicted	0%	67%	38%	27%	33%	0%				
Category Underpredicted	0%	0%	13%	47%	67%	0%				

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NRU=Neutral red uptake; RC=Registry of Cytotoxicity.
[1]The RC rat-only millimole regression is log LD_{50} (mmol/kg) = log IC_{50} (mM) x 0.439 + 0.621. Numbers in table represent numbers of substances.
[2]From **Table 3-1**.

The overall accuracy of the 3T3 NRU test method for correctly predicting GHS acute oral toxicity classification category of the 30 reference substances using the RC rat-only weight regression was 30% (see **Table 3-3**). *In vivo* toxicity was overpredicted for 33% and underpredicted for 37%. For this analysis, in terms of each GHS acute oral toxicity classification category:

- Zero (0%) of 5 substances with LD_{50} <5 mg/kg was correctly predicted
- One (20%) of 5 substances in the 5 < LD_{50} ≤50 mg/kg category was correctly predicted
- Three (60%) of 5 substances in the 50 < LD_{50} ≤300 mg/kg category were correctly predicted
- Three (60%) of 5 substances in the 300 < LD_{50} ≤2000 mg/kg category were correctly predicted.
- Two (40%) of the 5 substances in the 2000 < LD_{50} ≤5000 mg/kg category were correctly predicted.
- Zero (0%) of the 5 substances with LD_{50} >5000 mg/kg were correctly predicted.

Table 3-3 Prediction of GHS Acute Oral Toxicity Category by the 3T3 NRU Test Method Using the Recommended Reference Substances and the RC Rat-Only Weight Regression

Reference Rodent Oral LD_{50}[2] (mg/kg)	NRU- Predicted GHS Category (mg/kg)						Total	Accuracy	Toxicity Over-predicted	Toxicity Under-predicted
	LD_{50} <5	5< LD_{50} ≤50	50 < LD_{50} ≤300	300 < LD_{50} ≤2000	2000 < LD_{50} ≤5000	LD_{50} >5000				
LD_{50} < 5	0	0	3	2	0	0	5	0%	0%	100%
5< LD_{50} ≤50	0	1	1	3	0	0	5	20%	0%	80%
50 < LD_{50} ≤300	0	0	3	2	0	0	5	80%	0%	58%
300 < LD_{50} ≤2000	0	0	2	3	0	0	5	80%	20%	0%
2000 < LD_{50} ≤5000	0	0	0	3	2	0	5	0%	100%	0%
LD_{50} >5000	0	0	0	2	3	0	5	20%	80%	0%
Total	0	3	8	15	3	1	30	30%	33%	37%
Predictivity	0%	100%	33%	20%	40%	0%				
Category Overpredicted	0%	0%	44%	47%	0%	0%				
Category Underpredicted	0%	0%	22%	33%	60%	0%				

Abbreviations: GHS=Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005); NRU=Neutral red uptake; RC=Registry of Cytotoxicity.
[1]The RC rat-only weight regression is log LD_{50} (mgkg) = log IC_{50} (ug/mL) x 0.372 + 2.024. Numbers in table represent numbers of substances.
[2]From **Table 3-1**.

4.0 REFERENCES

EPA. 2002. Health Effects Test Guidelines OPPTS 870.1100 Acute Oral Toxicity. EPA 712–C–02–190. Washington, DC: U.S. Environmental Protection Agency.

Halle W. 1998. Toxizitätsprüfungen in Zellkulturen für eine Vorhersage der akuten Toxizität (LD_{50}) zur Einsparung von Tierversuchen. Life Sciences/ Lebenswissenschaften, Volume 1, 94 pp., Jülich: Forschungszentrum Jülich. English translation: Halle W. 2003. The Registry of Cytotoxicity: Toxicity testing in cell cultures to predict acute toxicity (LD50) and to reduce testing in animals. Altern Lab Anim 31:89-198.

ICCVAM. 2001. Guidance Document On Using *In Vitro* Data To Estimate In Vivo Starting Doses For Acute Toxicity. NIH Publication No. 01-4500. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/. [accessed 20 September 2006].

ICCVAM. 2006. In Vitro Acute Toxicity Test Methods Background Review Document (pending). [PLACEHOLDER]

OECD. 1996. Final Report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Methods. (Solna Report) Paris: Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 20 September 2006].

OECD. 2000. Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation. Environmental Health and Safety Monograph Series on Testing and Assessment No. 19. Paris France: OECD. Available: http://www.oecd.org/document/22/0,2340,en_2649_34377_1916054_1_1_1_1,00.html [accessed 20 September 2006]

OECD. 2001a. Guideline for Testing of Chemicals, 425, Acute Oral Toxicity – Up-and-

Down Procedure. Paris France: OECD. Available: http://www.oecd.org [accessed 20 September 2006].

OECD. 2001b. Guideline For Testing of Chemicals, 423, Acute Oral Toxicity – Acute Toxic Class Method. Paris France: OECD.

OECD. 2001c. OECD Series on Testing and Assessment, 33, Harmonized Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures. ENV/JM/MONO(2001)6. Paris France: OECD. Available: http://www.oecd.org [accessed 20 September 2006].

OECD. 2001d. Guideline for Testing of Chemicals, 420, Acute Oral Toxicity – Fixed Dose Method. Paris France: OECD. Available: http://www.oecd.org [accessed 20 September 2006].

OECD. 2001e. Guidance Document on Acute Oral Toxicity Testing. 24. Paris France: OECD. Available: http://www.oecd.org/ehs [accessed 9 October 2006].

OECD. 2002. Report of the Stockholm Conference on Validation and Regulatory Acceptance of New and Updated Methods in Hazard Assessment. Paris: Organisation for Economic Co-operation and Development. Available: http://www.oecd.org [accessed 20 September 2006].

PHS. 2002. Public Health Service (PHS) Policy on Humane Care and Use of Laboratory Animals. Office of Laboratory Animal Welfare. National Institutes of Health. Available: http://grants.nih.gov/grants/olaw/olaw.htm [accessed 25 September 2006].

Riddell RJ, Panacer DS, Wilde SM, Clothier RH, Balls M. 1986. The importance of exposure period and cell type in *in vitro* cytotoxicity tests. Altern Lab Anim 14:86-92.

SACATM. 2006. Meeting Minutes from the August 3, 2006 SACATM Working Group Meeting. http://ntp.niehs.nih.gov/files/SACATMAug06MinutesVF081506.pdf [accessed 16 October 2006].

Spielmann H, Genschow E, Liebsch M, Halle W. 1999. Determination of the starting dose for acute oral toxicity (LD_{50}) testing in the up and down procedure (UDP) from cytotoxicity data. Altern Lab Anim 27:957-966.

UN. 2005. Globally Harmonized System of Classification and Labelling of Chemicals (GHS), First Revised Edition. [ST/SG/AC.10/30/Rev.1]. United Nations, New York and Geneva. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html [accessed 20 September 2006].

APPENDIX A
PEER REVIEW PANEL REPORT

[This Page Intentionally Left Blank]

APPENDIX A1

PEER REVIEW PANEL REPORT: THE USE OF *IN VITRO* BASAL CYTOTOXICITY TEST METHODS FOR ESTIMATING STARTING DOSES FOR ACUTE ORAL SYSTEMIC TOXICITY TESTING

[This Page Intentionally Left Blank]

Peer Review Panel Report: The Use of *In Vitro* Basal Cytotoxicity Test Methods For Estimating Starting Doses For Acute Oral Systemic Toxicity Testing

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences
National Institutes of Health
U.S. Public Health Service
Department of Health and Human Services

June 2006

This document is available electronically at:

http://iccvam.niehs.nih.gov/methods/invidocs/panelrpt/ATpanelrpt.htm

TABLE OF CONTENTS

IN VITRO ACUTE TOXICITY PEER PANEL ROSTER

- **David H. Blakey**, D.Phil., Health Canada, Ottawa, Ontario, Canada
- **June Bradlaw**, Ph.D., International Foundation for Ethical Research (IFER), Rockville, Maryland
- **Robert Copeland**, Ph.D., Howard University College of Medicine, Washington, DC
- **Gianni Dal Negro**, D.V.M., Ph.D., GlaxoSmithKline Medicine Research Centre, Verona, Italy
- **Marion Ehrich**, Ph.D., RPh., D.A.B.T. Virginia-Maryland Regional College of Veterinary Medicine, Blacksburg, Virginia
- **Eugene Elmore**, Ph.D., University of California, Irvine, Irvine, California
- **Benjamin Gerson**, M.D., Thomas Jefferson University School of Medicine, Philadelphia, Pennsylvania
- **Michael Greene**, Ph.D., U.S. Consumer Product Safety Commission, Bethesda, MD
- **Janice Kuhn**, Ph.D., D.A.B.T., Stillmeadow Inc., Sugar Land, Texas
- **Daniel Marsman**, D.V.M., Ph.D., D.A.B.T., Procter & Gamble Company, Cincinnati, Ohio
- **Andrew Rowan**, Ph.D., Humane Society of the United States, Washington, DC
- **Hasso Seibert**, Ph.D., University Medical School Schleswig-Holstein, Kiel, Germany
- **Nigel Stallard**, Ph.D., The University of Warwick, Coventry, United Kingdom
- **Katherine Stitzel**, D.V.M., (Panel Chair), Consultant, West Chester, Ohio
- **Shinobu Wakuri**, MSc., Hatano Research Institute, Japan
- **Daniel Wilson**, Ph.D., D.A.B.T., The Dow Chemical Company, Midland, Michigan

PREFACE

This is an independent report of the *In Vitro* Acute Toxicity Peer Review Panel ("Panel") organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). The report summarizes discussions, conclusions, and recommendations of the public meeting of the Panel that was held at the National Institutes of Health in Bethesda, MD, on May 23, 2006. The ICCVAM and the Acute Toxicity Working Group (ATWG) will consider the Panel report, along with public comments, to prepare final test method recommendations for U.S. Federal agencies. ICCVAM test method recommendations will be forwarded to U.S. Federal agencies for consideration and action, in accordance with the ICCVAM Authorization Act of 2000 (P.L. 106-545).

NICEATM and the European Centre for the Validation of Alternative Methods (ECVAM) organized and conducted the NICEATM/ECVAM *In Vitro* Basal Cytotoxicity Validation Study. NICEATM, in coordination with the ATWG and ICCVAM, prepared a comprehensive draft background review document (BRD) reviewing the study. The draft BRD documents the procedures and results generated from the multi-phase study using the BALB/c 3T3 murine fibroblast (3T3) and normal human epidermal keratinocyte (NHK) neutral red uptake (NRU) test methods for the prediction of starting doses for acute oral toxicity test methods. The draft BRD was made publicly available on the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov) or from NICEATM on request.

NICEATM, in collaboration with the ATWG and ICCVAM, announced the independent Peer Panel review of the test methods in March 2005. Comments from the public and scientific community were solicited and provided to the Panel for their consideration (FR Notice Vol. 71, No. 54, pp. 14229-30, 3/21/06).

The Panel was charged with:

- Developing conclusions and recommendations regarding the usefulness and limitations of *in vitro* NRU basal cytotoxicity test methods using the 3T3 and NHK cells to estimate the rat oral acute LD_{50} for the purpose of determining the starting dose for *in vivo* acute oral toxicity test methods and thereby reducing animal use
- 'Peer reviewing' the NICEATM/ECVAM *In Vitro* Acute Toxicity Test Methods Draft BRD for completeness and for any errors or omissions
- Evaluating the information in the Draft BRD to determine the extent to which each of the applicable criteria for validation and acceptance of toxicological test methods (ICCVAM 2003[1]) have been appropriately

[1] ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. Research Triangle Park, NC:NIEHS. The guidelines can be obtained at: http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm

addressed (validation[2] of a new test method is a prerequisite for it to be considered for regulatory decision-making)

- Considering the ICCVAM draft test method recommendations for these test methods (i.e., the proposed test method uses, the proposed recommended standardized protocols, and the proposed test method performance standards) and comment on whether the recommendations are supported by the information provided in the Draft BRD

During the public meeting on May 23, 2006, the Panel discussed the current validation status of the *in vitro* test methods. The Panel also provided formal comment on the Draft BRD and made recommendations for revisions to the Draft BRD. The Panel also provided formal comment on the ICCVAM recommendations for test method use, future studies, test method performance standards, and the cytotoxicity protocols. In addition, the public were provided time at the public meeting to comment on the Draft BRD. The Panel then provided final endorsement regarding the validation status of the test methods.

The Panel gratefully acknowledges the efforts of the NICEATM staff in coordinating the peer review logistics and accommodations and in the preparation of the Draft BRD and various other materials for the review.

[2] Validation is the process by which the reliability and accuracy of a test method are established for a specific purpose (ICCVAM 2003).

EXECUTIVE SUMMARY

Introduction

This report describes the conclusions and recommendations of the *In Vitro* Acute Toxicity Peer Panel ("Panel") regarding the validation status of the BALB/c 3T3 murine fibroblast (3T3) and normal human epidermal keratinocyte (NHK) *in vitro* neutral red uptake (NRU) basal cytotoxicity test methods (hereafter designated as NRU test methods) and the ability to use these test methods to estimate starting doses for acute oral systemic toxicity tests. The Panel accepts the sections of the Draft Background Review Document for *In Vitro* Acute Toxicity Test Methods (BRD) for which it had no comments and recommendations as adequate and acceptably accurate.

Panel Recommendations for the BRD

The Panel stated that, in general, the information presented in the Draft BRD was sufficient for its purpose. Exceptions are noted within the body of the Panel report. The Panel concluded that the objectives of the validation study were appropriate, and agreed that the applicable validation criteria were adequately addressed in the Draft BRD for using these *in vitro* test methods to determine starting doses for acute oral systemic toxicity tests.

The Panel made numerous recommendations for additional explanations (e.g., provide the rationale for using serum that is not heat-inactivated) and clarifications (e.g., provide additional details for using the GraphPad PRISM® software to calculate IC_{50} values) to the Draft BRD that will not require additional statistical analyses. Some recommendations included presentation of the existing data in other formats (e.g., using the relative IC_{50} ratios between the reference substances and the positive control [at the level of the individual laboratory] to compare similar substances across test methods), or additional analyses (e.g., determine the usefulness of the test methods to estimate starting doses for the Fixed Dose Procedure [FDP] acute toxicity test method).

The Panel concluded that several confounding factors were not addressed in the selection or evaluation of test substances but should be. They recommended that the octanol:water coefficients and the surface-active potential (to the extent possible) for the 72 reference substances should be characterized and incorporated into the assessment of accuracy. The Panel also recommended that protein binding should also be taken into account in the data analyses (i.e., to the extent possible, the free fraction in serum corresponding to the LD_{50} should be considered). Another potential confounder was the attempt to select chemicals to prevent the entire set of reference substances from having proportionally more *outlier* substances than the Registry of Cytotoxicity (RC) linear regression.

In the evaluation of test method accuracy, substances with neurotoxic and cardiotoxic mechanisms, and those that interfere with energy utilization or that alkylate cellular macromolecules were excluded. Such substances were excluded because it was expected that these mechanisms of action could not be detected by the NRU test methods. The Panel disagreed with their exclusion because interference with energy metabolism and alkylation of proteins and deoxyribonucleic acid (DNA) represent important mechanisms of cytotoxicity

that should be detected by these two test methods. Additionally, there was consensus among the Panel members that the available data on the mechanism of acute *in vivo* toxicity were not sufficient to justify the exclusion of substances based on mechanism and/or possible involvement of biotransformation reactions. However, the Panel recommended that the properties (e.g., metabolism, receptors, transporters) of the cell types that are important for basal cytotoxicity be better characterized. Despite the fact that there was no significant difference between rat and mouse LD_{50} data from the RC, the Panel indicated that the separation of such data (in developing *in vitro-in vivo* regressions) is useful because it decreases the biological variability associated with species differences.

Although the Panel recommended additional analyses for the evaluation of intra- and inter-laboratory reproducibility (i.e., the comparison of ratios of the maxima and minima mean laboratory IC_{50} values), the Panel agreed that these would not change the conclusion that the NHK NRU test method was more reproducible than the 3T3 version. The Panel suggested that an explanation for the difference in interlaboratory reproducibility be provided.

The Panel recommended that the analyses to determine the reduction of animal use consider prevalence (i.e., the distribution of the universe of substances that are likely to be tested within each hazard classification). The Panel also recommended that animal reduction/refinement be evaluated for the use of the NRU test methods to determine the starting dose for the FDP.

The Panel suggested that costs for equipment and working time needed to perform the NRU test methods and a cost-benefit analysis, including information on the reduction of the number of animals used, should be included in the Draft BRD. The time needed to prescreen NHK culture medium should also be included.

Validation Status of the NRU Test Methods

The Panel agreed that the applicable validation criteria have been adequately addressed for using these *in vitro* test methods in a weight-of-evidence approach to determine the starting dose for acute oral *in vivo* toxicity protocols. However, the Panel was aware that validation of the two NRU test methods was carried out not only to determine if they could be used to set starting doses for *in vivo* acute toxicity studies, but also to determine the extent to which the tests could be useful step in an *in vitro* tiered testing strategy for acute toxicity. The Panel agreed the validation study showed that neither of the two NRU test methods evaluated could be used as a stand-alone replacement for the *in vivo* tests even considering the variability of the latter. The Panel encouraged future work to develop a tiered testing strategy that includes basal cytotoxicity as part of the overall strategy.

Review of the Draft Interagency Coordinating Committee on Validation of Alternative Methods (ICCVAM) Recommendations for Test Method Use

The Panel agreed that although neither of the NRU test methods can be used as alternatives for the *in vivo* acute oral toxicity test for the purposes of hazard classification, the test

methods may be useful in a weight-of-evidence approach to determine the starting dose for acute oral *in vivo* toxicity protocols. The Panel agreed that the NRU test methods be considered before animals are used if there was no other stronger weight-of-evidence information on which to base a starting dose.

The Panel disagreed that the NRU test methods were not appropriate for substances that interfere with energy utilization or alkylation of proteins and other macromolecules and with using the revised RC regression that excluded chemicals based on mechanism of action. However, the Panel agreed with using the RC rat-only regression to estimate the LD_{50} from IC_{50} data and agreed that a regression based on weight rather than molar units would be useful for situations where the molar weight of the test substance is unknown. In situations where the molecular weight of a test substance is known, the molar regression should be used.

The Panel agreed that other *in vitro* basal cytotoxicity test methods are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD_{50} value, respectively) should be demonstrated to meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods.

Some Panel members agreed that the 3T3 NRU, based on relative ease of performance and cost, should be recommended for general use, but cautioned that one test method should not be preferred over the other. One Panel member noted that it is important to remember that hazard assessment relates to the safety of humans, not rats. The NHK NRU IC_{50} data had a higher correlation with human LC_{50} values (R^2=0.62) than did rodent 3T3 NRU IC_{50} data (R^2=0.51) and a higher correlation than did rodent LD_{50} data with human LC_{50} values (R^2=0.56) (Casati et al. 2005).

Review of the Draft ICCVAM Recommendations for Future Studies

The Panel indicated that high quality comparative *in vitro* basal cytotoxicity data should be collected in tandem with *in vivo* rat acute oral toxicity test results to further evaluate the use of the these test methods for predicting the starting dose for acute oral toxicity tests. However, no Panel member recommended that *in vivo* testing be conducted solely to collect data to further assess the usefulness of the NRU test.

The Panel agreed that additional *in vitro* tests and other methods necessary to achieve accurate acute oral hazard classification should be investigated. The Panel also agreed that the *in vivo* database of reference substances used in the validation study be used to evaluate the utility of other non-animal approaches to estimate starting doses for rat acute oral toxicity tests.

The Panel agreed that standardized procedures to collect information pertinent to an understanding of the mechanisms of lethality should be included, to the extent possible, in future rat acute oral toxicity studies. Such information will likely be necessary to support the further development of predictive mechanism-based *in vitro* test methods. The Panel recommended that ICCVAM consider convening a working group to explore mechanisms of

action of acute toxicity and approaches for acquiring additional information on acute toxic mechanisms during acute toxicity testing.

The Panel agreed that an expanded list of reference substances with estimated rat LD_{50} values substantiated by high quality *in vivo* data should be developed for use in future *in vitro* test method development and validation studies and that there should be a concerted effort to obtain higher quality proprietary data from regulated industries.

Review of the Draft Performance Standards for *In Vitro* Acute Toxicity Test Methods and Draft Recommended Test Method Protocols

The Panel agreed that the available data from the validation study appeared to support the validity of the recommended performance standards for the two NRU test methods. The usefulness and limitations were well covered. Although the two NRU test methods may be useful, there would be cause for concern if use of the test methods were made compulsory for regulatory purposes as other information such as structure-property relationships, when available, could provide better estimates of starting doses for acute toxicity studies.

The Panel identified several aspects of the performance standards that should be clarified. Specifically, the Panel recommended that more thorough explanations and more detail for test method procedures should be added to the recommended test method protocols but that an effort should be made to streamline them, where possible, to assure easy use and transferability. Clarification of solubility procedures for the determination of test substances should be provided since the variability between laboratories in the selection of solvent indicates a possible flaw in the solvent determination procedure. The Panel also suggested including other methods for calculating the IC_{50} values and a recommendation for task-specific training for laboratory technicians.

1.0 Introduction And Rationale for the Use of *In Vitro* Neutral Red Uptake (NRU) Cytotoxicity Test Methods to Predict Starting Doses for *In Vivo* Acute Oral Systemic Toxicity Testing

This section of the Draft *In Vitro* Acute Toxicity Test Methods Background Review Document (BRD) provided valuable historical background on the use of *in vitro* NRU test methods to predict starting doses for *in vivo* acute oral systemic toxicity. The objectives of the validation study were valid. The introduction acknowledged that *in vitro* cytotoxicity could not replace the Up-and-Down Procedure (UDP) or the Acute Toxic Class method (ATC) acute oral toxicity tests in animals. Furthermore, these *in vitro* tests would not be appropriate substitutes for any of the other standard acute toxicity tests. The Draft BRD recommended that *in vitro* cytotoxicity testing be part of a weight-of-evidence approach to determining the starting dose for *in vivo* acute oral systemic toxicity testing.

1.1 Background and Rationale for the Use of *In Vitro* Cytotoxicity Test Methods to Predict Starting Doses for *In Vivo* Acute Oral Systemic Toxicity Tests

This section briefly mentioned the concept of using the predicted LD_{50} value as a starting dose for acute oral toxicity to reduce the number of animals. This was first discussed at a European Centre for the Validation of Alternative Methods (ECVAM) workshop in 1996 (Seibert et al. 1996). The Panel suggested that this section also include the other major conclusions and recommendations of that workshop. The 1996 ECVAM workshop arrived at a general consensus, that

- Testing for basal cytotoxicity is not sufficient for prediction of acute systemic toxicity.
- Biokinetic factors must be considered before performing *in vitro*/*in vivo* comparisons, in order to make the *in vivo* and *in vitro* data more comparable and the resulting comparison more meaningful.

The Panel also recommended including information from an international project supported by the Commission of the European Communities. The project was performed in 1992 and 1993 by the Fund for Replacement of Animals in Medical Experiments (FRAME); Institute of Toxicology, Kiel, Germany; University of Nottingham, United Kingdom (UK); and Gesellschaft für Strahlen- und Umweltforschung (Society for Radiological and Environmental Research, for which the name changed to Forschungszentrum für Umwelt und Gesundheit [Center for Environmental and Health Research]), Neuherberg, Germany. The report, *An International Evaluation of Selected in Vitro Toxicity Test Systems for Predicting Acute Systemic Toxicity* (Fentem et al. 1993), contains results on the *in vitro* cytotoxicity of 42 substances determined with a 3T3 NRU test method and several other *in vitro* systems. Many of the substances tested are identical to those tested in the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)/ECVAM validation study. Furthermore, the report contains statistical analyses of correlations between rodent LD_{50} values and *in vitro* IC_{50} values, and evaluations of the accuracy of the *in vitro* methods for predicting LD_{50} values and acute toxicity categories, respectively.

The Registry of Cytotoxicity (RC) is a registry of lethality and IC_{50} values. The Panel agreed that this database is important and that increasing the numbers of chemicals in this database would be of value. However, IC_{50} values do not indicate the steepness of slope for the cytotoxicity concentration response relationship nor the number of points the value is based on. Furthermore, the RC used many endpoints for cytotoxicity, some of which may be reversible (e.g., cell detachment, effects on cell proliferation). These deficiencies must be mentioned.

The stepwise approach for the validation study was a good approach because it allowed for the review of intermediate progress.

1.2 Regulatory Rationale and Applicability for the Use of *In Vitro* Cytotoxicity Test Methods to Predict Starting Doses for Acute Oral Systemic Toxicity Testing

1.2.1 *Current Regulatory Testing Requirements for Acute Systemic Toxicity*

This section provided a great deal of detail regarding the context of the regulatory requirements for acute oral toxicity assays.

1.2.2 *Intended Regulatory Uses for the In Vitro Cytotoxicity Test Methods*

This section should clarify that the NRU cytotoxicity test methods are to be used in a weight-of-evidence approach to determining the starting dose for acute oral systemic toxicity assays. The default starting dose is usually used when there is no information upon which to base a starting dose (e.g., no toxicity information from chemicals with similar structure, etc.).

The Draft BRD indicated that the NRU cytotoxicity test methods could not be used to determine the starting dose for the Fixed Dose Procedure (FDP) because it is not possible to predict a dose that leads to non-fatal toxicity (the TD_{50}). The Panel suggested the TD_{50} and IC_{50} are highly correlated, so that, given TD_{50} data, a regression model for prediction of TD_{50} from IC_{50} could be constructed. Even in the absence of TD_{50} data, a simple procedure such as assuming that the FDP starting dose is two doses below the estimated LD_{50} would be worth investigating. The studies of one Panel member, who has compared IC_{50} values for growth inhibition and mitochondrial function of various epithelial cell lines from normal human tissues, showed that adverse events in clinical studies were observed only after plasma levels exceeded the *in vitro* IC_{50} levels by about one log or more.

1.2.3 *Similarities and Differences in the Endpoints of the In Vitro Cytotoxicity Test Methods and In Vivo Acute Oral Toxicity Test Methods*

Animal death and death of cells in culture may or may not have similarities at the cellular level. As noted in the Draft BRD, extrapolation to the whole organism may involve more than just cellular death.

The Draft BRD recognized the ability of normal human epidermal keratinocytes (NHK) to metabolize some xenobiotic substances. The fact that BALB/c mouse fibroblast 3T3 (3T3) cells and NHK cells responded differently to several of the reference substances tested could result from differences in doubling times between the two cell lines. It also could result from detoxification mechanisms or metabolites generated in the NHK cells. The use of serum can

complicate the issue of determining and/or identifying mechanism of toxicity. The 3T3 cell culture system included serum, while the NHK cell culture system did not. Mechanistic differences in cell type are recognized for toxicants that act at particular receptors.

Toxin should be used to refer to a biological product. Since the NICEATM/ECVAM validation study tested pure chemicals, the term *toxicant* should be used.

1.2.4 *Use of In Vitro Cytotoxicity Test Methods in the Overall Strategy of Hazard Assessment*

The Draft BRD indicated that the RC millimole regression cannot be used with mixtures and unknown substances because the equation requires molecular weight information for the mole units. The new regression formula (developed in Section 6) based on gram units should be described in this section, too. The new regression formula would be applicable to mixtures and unknown substances.

1.3 Scientific Basis for the *In Vitro* NRU Test Methods

1.3.1 *Purpose and Mechanistic Basis of the In Vitro NRU Test Methods*

The Draft BRD should clarify the extent to which Borenfreund and Puerner (1985) relied on morphology to determine the maximal tolerated dose.

1.3.2 *Similarities and Differences in the Modes/Mechanisms of Action for the In Vitro NRU Test Methods Compared with the Species of Interest*

This section well delineated the differences between the cell types.

1.3.3 *Range of Substances Amenable to the In Vitro NRU Test Methods*

This section of the Draft BRD appropriately identified problems concerning substances with specific toxicity mechanisms, those that were insoluble or volatile, the presence of serum, lysosomal sequestration, and red color. It should be noted that other colored compounds may present a problem as well.

2.0 Test Method Protocol Components of the 3T3 and NHK *In Vitro* NRU Test Methods

The information presented in Section 2 of the Draft BRD appeared to be sufficient. There was a great deal of detail regarding the equipment, methods, and procedures required for implementation of the proposed 3T3 and NHK NRU test methods.

The Guidance Document (ICCVAM 2001b) recommendations were good. This section should explain why it is important to have an exposure period of at least the duration of one cell cycle.

2.1 Overview of the 3T3 and NHK NRU Test Methods

This section of the Draft BRD noted the similarities and differences of the 3T3 and NHK NRU cytotoxicity test methods. The similarities included preparation of reference substances

and the positive control, cell culture environmental conditions, determination of test substance solubility, 96-well plate configuration, 48 hour exposures, microscopic evaluation, NRU measurement as % of control with concentration in μg/mL, and data analysis. The 3T3 and NHK NRU differed in conditions for cell propagation, cell growth media, and application of reference substances (volume). The Panel noted that the IC_{50} values obtained during the study are only valid under the conditions used in the conduct of the test methods.

2.1.1 *The 3T3 NRU Test Method*

The Panel noted that the serum for the 3T3 NRU test method was not heat-inactivated. Serum that is not heat-inactivated can contain enzymes (i.e., esterases) that transform certain chemicals. The Draft BRD should explain the rationale for using serum that is not heat-inactivated. Of the 21 substances deleted from the accuracy analyses (Table 6.3 of the Draft BRD), one Panel member noted that eight substances (atropine, carbamazepine, dichlorvos, disulfoton, fenpropathrin, parathion, physostigmine, procainamide) had structures that could have been biotransformed by serum enzymes.

The Draft BRD should also discuss the rationale for the restriction of the use of the 3T3 cells to less than 18 passages after thawing.

2.1.2 *The NHK NRU Test Method*

Keratinocytes were not subcultured beyond the second passage, which is not unusual for primary cells. The Draft BRD should acknowledge that the use of different lots of NHK cells by an investigator might increase variability.

2.1.3 Measurement of NRU for both 3T3 and NHK Test Methods

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2 Descriptions and Rationales of the 3T3 and NHK NRU Test Methods

The Draft BRD mentioned that there were problems concerning the growth of both the 3T3 and NHK cells. Since the growth rate can be very important for the results of the cytotoxicity test methods, the Draft BRD should report the doubling times after seeding the cells in 96-well plates and during exposure.

2.2.1 *Materials, Equipment, and Supplies*

Materials and equipment were listed in this section. There was no information regarding the maximum absorbance required of the plate reader; this must be provided as many spectrophotometers following Beer's Law can only read a maximum optical density (OD) of ~ 3.

2.2.2 *Reference Substance Concentrations/Dose Selection*

A commercial medium (keratinocyte basal medium [KBM®] supplied by Clonetics®) was used for culturing the NHK cells. There was no specific information on the composition of this medium. The exact composition of the medium should be specified, especially, whether sera are included, and, if so, the types and concentrations. Without this information, it is

impossible to judge whether differences in medium composition may contribute to the differing results of the test methods for several of the test substances.

2.2.3 *NRU Endpoints Measured*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2.4 *Duration of Reference Substance Exposure*

The 48-hour duration of exposure was justified in this section. The differences between *in vitro* cytotoxicity at 24- and 72-hour exposures were noted. As part of future research, it might be of interest to extend the duration of exposure to 96 hours to parallel the 4-day exposure used in animal test protocols. On the other hand, a time course may be important. Recovery and cell growth would suggest that an agent's IC_{50} could change at 72 or 96 hours relative to that at 48 hours. If recovery occurs, then lethality would require a higher dose. Perhaps two time points as used by Elmore (2001, 2002) would be useful. These studies used three days and five days for exposure and noted differences in the IC_{50} values. These time points were chosen to facilitate detection of growth inhibition. Increasing toxicity at five days suggested the agent was more toxic while decreasing toxicity suggested recovery of the cells.

2.2.5 *Known Limits of Use*

This section of the Draft BRD contained caveats on solubility, volatility, and pharmacokinetics, noting that the latter was not addressed. The organ-specific section contained a 5-step *in vitro* test method. The value of including this organ-specific section was unclear since it did not refer to the use of organ-specific cells. The organ-specific section was more concerned with metabolism, energy production, and disruption of epithelial barriers.

Another limitation of use of the *in vitro* test methods is for substances that etch plastics and those that film out (i.e., form a film on the medium surface or plastic well wall). Substances that etch plastics can be detected by looking for the presence of etched rings in the 96-well plates after exposure. Some substances that film out in medium may etch plastic. Additionally, substances that film out decrease the concentration delivered to the cells. Such substances can be identified by the changes produced in the meniscus of the medium or by the presence of a film where the surface of the medium was in the well.

2.2.6 *Nature of Response Assessed*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.2.7 *Appropriate Vehicle, Positive, and Negative Controls*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

598 2.2.8 *Acceptable Ranges of Control Responses*
599 The Draft BRD should explain why vehicle control (VC) ODs were lower during Phase II
600 and Phase III testing. Higher viability appeared to correlate with high absorbance. The VC
601 OD ranges of each laboratory should be described so that the stability of cell growth
602 conditions in each laboratory can be evaluated.
603
604 The doubling time of each cell type (for each laboratory) should be described in this section.
605
606 2.2.9 *Nature of Experimental Data Collected*
607 Since the Study Director decided whether to remove outliers at 99% level, the Study Director
608 must be an expert in theory and practice of cell culture.
609
610 2.2.10 *Type of Media for Data Storage*
611 The Panel found that the Draft BRD discussion and evaluation in this section was
612 appropriate.
613
614 2.2.11 *Measures of Variability*
615 The Panel found that the Draft BRD discussion and evaluation in this section was
616 appropriate.
617
618 2.2.12 *Methods for Analyzing NRU Data*
619 The Panel found that the Draft BRD discussion and evaluation in this section was
620 appropriate.
621
622 2.2.13 *Decision Criteria for Classification of Reference Substances*
623 The Panel found that the Draft BRD discussion and evaluation in this section was
624 appropriate.
625
626 2.2.14 *Information and Data Included in the Test Report*
627 The Panel found that the Draft BRD discussion and evaluation in this section was
628 appropriate.
629
630 2.3 Basis for Selection of the *In Vitro* NRU Cytotoxicity Test Methods
631
632 The selection of NRU cytotoxicity test methods was derived from the Report of the
633 International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity
634 (ICCVAM 2001a). Workshop participants evaluated several *in vitro* initiatives to evaluate
635 the prediction of systemic toxicity from *in vitro* toxicity. Workshop participants concluded
636 that there were no differences between species sources or between continuous cell lines and
637 primary cells.
638
639 2.3.1 *Guidance Document Rationale for Selection of In Vitro NRU Cytotoxicity Test*
640 *Methods*
641 The Panel found that the Draft BRD discussion and evaluation in this section was
642 appropriate.
643
644

644 2.3.2 *Guidance Document Rationale for Selection of Cell Types*
645 ICCVAM wanted rodent cells used in a cytotoxicity test method because LD_{50} data is
646 obtained with rodents. Cell lines rather than primary cultures would hasten generation of an
647 *in vitro* database. Highly differentiated cells were not used and neither were metabolically
648 active cells such as liver.
649
650 2.4 Proprietary Components of the *In Vitro* NRU Cytotoxicity Test Methods
651
652 Proprietary cells and media were used for the NHK NRU method (Clonetics®).
653
654 2.5 Basis for Number of Replicate and Repeat Experiments for the 3T3 and NHK NRU
655 Test Methods
656
657 The Panel found that the Draft BRD discussion and evaluation in this section was
658 appropriate.
659
660 2.6 Basis for Modifications to the 3T3 and NHK NRU Test Method Protocols
661
662 The Panel recommended that the OD of the positive control be included in Table 2-2 of the
663 Draft BRD. The VC OD range was eventually deleted as a test acceptance criterion.
664
665 The Panel asked whether something other than mechanism of action contributed to the
666 unusual concentration-response curves for aminopterin and colchicine. The Draft BRD
667 should identify those substances for which the IC_{50} was calculated using only one point
668 between 0 and 100% when a substance had a steep concentration-response curve. The Panel
669 preferred that there be three points between 10 and 90% viability
670
671 2.6.1 *Phase Ia: Laboratory Evaluation Phase*
672 The ring of dead NHK cells was produced by the use of the plate inversion technique for
673 removing the cell culture medium prior to refeeding the cells. Such a technique leaves
674 residual media around the edges of each well. The ring of dead cells can be avoided by
675 aspirating the medium from the wells prior to refeeding. Aspiration also obviates the need to
676 prepare chemicals as a 2X dilution. A 1X chemical solution (or vehicle control) can be added
677 to the cells immediately after aspiration to avoid drying of the cells.
678
679 2.6.2 *Phase Ib: Laboratory Evaluation Phase*
680 The Panel found that the Draft BRD discussion and evaluation in this section was
681 appropriate.
682
683 2.6.3 *Phase II: Laboratory Qualification Phase*
684 The approach for handling of volatile agents by covering the 96-well plates with plastic film
685 was appropriate. The Panel recommended that oil not be used to cover the culture media
686 surface because agents that bind to lipids can bind to the oil, which reduces their effective
687 concentration.
688

Prism® software calculations for IC_{50} using Hillslope and midpoints may under- or overestimate the IC_{50} depending on the inclusion of nontoxic concentrations for which viability is >100%, highest test concentrations that produce less than complete toxicity (i.e., viability >0%), or concentration-response curves for which the lowest nontoxic concentration produced <100% viability. The Panel was not satisfied with the current explanation for the IC_{50} calculation.

2.6.4 *Phase III: Laboratory Testing Phase*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.7 Differences in 3T3 and NHK NRU Test Method Protocols and the *Guidance Document* Standard Protocols

2.7.1 *Optimization of the Guidance Document Protocols Prior to Initiation of the Study*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.7.2 *Optimization of the Guidance Document Protocols During the Study*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.8 Overview of the Solubility Protocol

A complex flow chart for the solvent selection for each test substance was provided.

2.9 Components of the Solubility Protocol

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.9.1 *Medium, Supplies, and Equipment Required*

The Panel suggested that the visual solubility determination be performed using a microscope.

2.9.2 *Data Collection*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate

2.9.3 *Variability in Solubility Measurement*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

2.9.4 *Solubility and the 3T3 and NHK NRU Test Methods*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate

734 2.9.5 *Methods for Analyzing Solubility Data*
735 The Panel found that the Draft BRD discussion and evaluation in this section was appropriate
736
737 2.10 Basis of the Solubility Protocol
738
739 The Panel had no comments on this section, although the comments on the protocol itself are
740 addressed below.
741
742 2.10.1 *Initial Solubility Protocol Development*
743 The Draft BRD noted that sometimes BioReliance and the cytotoxicity testing laboratories
744 did not get the same solubility results and additional explanation as to why this occurred
745 would be useful. However, as a whole, solubility was not a major issue.
746
747 2.10.2 *Basis for Modification of the Phase II Protocol*
748 The Panel found that the Draft BRD discussion and evaluation in this section was
749 appropriate.
750
751 2.11 Summary
752
753 The Panel found that the Draft BRD discussion and evaluation in this section was
754 appropriate.
755
756 **3.0 Reference Substances Used for Validation of the 3T3 And NHK NRU Test**
757 **Methods**
758
759 3.1 Rationale for the Reference Substances Selected for Testing
760
761 The selection of test chemicals, the determination of reference *in vivo* data, as well as test
762 method standardization and validation appeared to be well described, and generally of high
763 quality. A wide range of substances, belonging to many chemical classes, with varying
764 physical properties, and different mechanisms of toxicity were included. The list included
765 pharmaceuticals, pesticides, solvents and a number of metal-containing molecules; however,
766 there were no polycyclic aromatic hydrocarbons, catalysts, simple aldehydes, ketones,
767 biocides, cosmetic ingredients, mixtures/formulations, plant toxins, or other natural
768 compounds. The molecular structures were not provided and should be.
769
770 The adequacy of the range of reference substances and their mechanisms of oral toxicity was
771 difficult to judge because there is often very limited knowledge about their mechanisms of
772 action. The overall poor characterization of modes or mechanisms of action of acute oral
773 toxicity *in vivo* makes it difficult to strategically select reference substances for broad acute
774 toxicity validation of *in vitro* methods. However, since the NRU methods are expected to
775 detect basal cytotoxicity, the selected substances should be sufficient to evaluate reliability
776 and accuracy. Specifically, the Draft BRD provided little information about the 72 reference
777 substances to indicate that specific modes of action of acute systemic toxicity had been
778 robustly explored.
779

The standardized methodology for acute toxicity protocols (i.e., the traditional LD_{50} or the reduced UDP procedure), which include only the most rudimentary collection of endpoints, makes no attempt to characterize even the simplest modes of action of a test substance. As such, the overall poor characterization of these reference substances for modes or mechanisms of action of acute oral toxicity *in vivo* made it difficult to strategically select reference substances for broad acute toxicity validation of *in vitro* methods.

Within this context, there may be some limited value in adding data from additional substances to improve precision. Inclusion of substances at the extremes of the GHS toxicity categories may be helpful.

3.1.1 *Reference Substance Selection Criteria*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.1.2 *Candidate Reference Substances*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.1.3 *Selection of Reference Substances for Testing*

The selection of reference substances for evaluating the reliability and the accuracy of the NRU cytotoxicity test methods was well planned and executed, arriving at a broad and fairly complete selection of model chemicals. However, many test substances in the regulatory testing realm are mixtures. It would have been useful to include some mixtures similar to common pesticide or household product formulations.

Also regarding the selection of reference substances used to determine the accuracy of the 3T3 and NHK test methods, there was an attempt to maintain the same proportion of "outliers" as was present in the RC. However, the total percentage of RC outliers in the set of reference substances (38%) was greater than the total percentage of outliers in the RC (27%). This should be highlighted and addressed as a potential confounder. Conversely, there was some concern that the potential for bias may exist if chemicals were pre-selected based on best fit to a regression line plotting cytotoxicity versus *in vivo* LD_{50} to evaluate *in vitro* test methods to estimate the acute oral LD_{50}. This bias likely predisposed the results to overprediction of the value of the NRU test methods for predicting random source chemicals. This potential bias needs to be discussed.

3.2 Rationale for the Number of Reference Substances Selected

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.3 Characteristics of the Selected Reference Substances

3.3.1 *Source Databases Represented by the Selected Reference Substances*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.3.2 *Chemical Classes Represented by the Selected Reference Substances*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.3.3 *Product/Use Classes Represented by the Selected Reference Substances*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.3.4 *Toxicological Characteristics of the Selected Reference Substances*

Several confounding factors were addressed in the selection or evaluation of the reference substances (e.g., the octanol:water partition coefficient and the surface-active potential). These should be characterized and this information should be incorporated into the assessment.

Surface active molecules, in particular those that can partition at the oil-water interface, can significantly influence absorption, toxicity, and interactions with other molecules, and may enhance or diminish the predictive capacity of an *in vitro* test method. Test substance concentration and inherent toxic potential may be heavily influenced by molecular charge and surface activity.

Another example of a physical-chemical feature that can represent a confounding factor is given by the cationic amphiphilic molecules that contain a hydrophobic ring structure and a hydrophilic side chain with a charged cationic amine group. This chemical structure enables the substance to penetrate the cell membranes very rapidly and to interfere with phospholipid metabolism, causing phospholipidosis. This issue needs to be addressed.

3.3.5 *Selection of Reference Substances for Testing in Validation Study Phases Ib and II*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

3.3.6 *Unsuitable and Challenging Reference Substances*

The cytotoxicity endpoint for the test method is based on uptake of neutral red into lysosomes. The Draft BRD did not mention whether any of the reference substances cause lysosomal swelling, which could cause artifacts.

3.4 Reference Substance Procurement, Coding, and Distribution

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

868 3.5 Reference Substances Recommended by the *Guidance Document* (ICCVAM
869 2001b)
870
871 The Panel found that the Draft BRD discussion and evaluation in this section was
872 appropriate.
873
874 3.6 Summary
875
876 To the extent possible, characterization of the metabolic profiles of the reference substances
877 should be added.
878
879 **4.0 *In Vivo* Rodent Toxicity Reference Values Used to Assess the Accuracy of the**
880 **3T3 and NHK NRU Test Methods**
881
882 This section described the problems that arise in finding and using rodent LD_{50} values taken
883 from the published literature. These problems have been well known for decades (e.g., a
884 review by Morrison et al. 1968) and little has improved since then as indicated by the lack of
885 data collected under Good Laboratory Practice (GLP) guidelines. Given the shortcomings of
886 the existing data, the information provided was adequate and revisions are unlikely to lead to
887 any significant improvement.
888
889 The mechanisms of oral toxicity of the reference substances were difficult to determine
890 because LD_{50} values are so rarely accompanied by more detailed information concerning the
891 actual lesions observed and the reason for the animals' deaths. The overall poor
892 characterization of modes or mechanisms of acute toxicity resulted in some difficulty in
893 developing more sophisticated comparisons of *in vitro* and *in vivo* data.
894
895 4.1 Methods Used to Determine *In Vivo* Rodent Toxicity Reference Values
896
897 4.1.1 *Identification of Candidate In Vivo Rodent Toxicity Reference Data*
898 The selection of reference *in vivo* data was well described. A wide range of databases was
899 searched and a comprehensive set of *in vivo* LD_{50} identified. In general, the actual data did
900 not conform to modern standards of toxicity testing, hence their quality would be difficult to
901 determine (99% - 452 of 459 LD_{50} values would have to be eliminated if a GLP requirement
902 were to be mandated).
903
904 4.1.2 *Criteria Used to Select Candidate In Vivo Rodent Toxicity Data for Determination*
905 *of Reference Values*
906 The Panel found that the Draft BRD discussion and evaluation in this section was
907 appropriate.
908
909 4.2 Final *In Vivo* Rodent Toxicity Reference Values
910
911 The Panel found that the Draft BRD discussion and evaluation in this section was
912 appropriate.
913

914 4.3 Relevant Toxicity Information for Humans
915
916 The Panel found that the Draft BRD discussion and evaluation in this section was
917 appropriate.
918
919 4.4 Accuracy and Reliability of the *In Vivo* Rodent Toxicity Reference Values
920
921 Because many of the 72 reference substances had multiple LD_{50} values in the literature, these
922 values had to be transformed to a single reference value for each chemical. The mean
923 maximum:minimum values for those chemicals that had multiple LD_{50} values showed a
924 tendency to decline as the toxicity decreased (See Table 4.4 of the Draft BRD). This may
925 simply reflect the fact that inherent biological variability has a greater impact at low LD_{50}
926 values than at high.
927
928 4.5 Summary
929
930 There was a general consensus that adequate data have been generated to draw conclusions
931 about the accuracy and validity of the methods. The majority of the most relevant *in vivo* data
932 from the available literature were collected to compare the two *in vitro* tests with *in vivo*
933 acute toxicity in rodents.
934
935 **5.0 3T3 and NHK NRU Test Method Data and Results**
936
937 In general, the results section adequately presented the data and results. The statistical
938 methods adopted provide a good quality analysis. However, several outcomes (indicated in
939 the following subsections) were not adequately addressed.
940
941 5.1 3T3 and NHK NRU Test Method Protocols
942
943 The Panel found that the Draft BRD discussion and evaluation in this section was
944 appropriate.
945
946 5.2 Data Obtained to Evaluate Accuracy and Reliability
947
948 5.2.1 *Positive Control (PC) Data*
949 The Draft BRD should explain the considerably higher sensitivity of NHK cells to the
950 positive control (sodium lauryl sulfate [SLS]).
951
952 5.2.2 *Reference Substance Data*
953 Consistently, carbon tetrachloride could not be tested in the 3T3 and NHK NRU test
954 methods. The reason that this chemical could not be tested should be addressed. Several
955 additional reference substances could not be adequately tested by one or two of the three
956 laboratories, although they had used the same cell types and harmonized protocols. The
957 reason(s) for these differences between the laboratories should be discussed.
958
959

5.3 Statistical Approaches to the Evaluation of 3T3 and NHK NRU Data

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.4 Summary of Results

Further discussion exploring the biological significance of and possible reasons for the differences in sensitivity and selectivity between the two cell lines is needed; this may be useful for selecting the appropriate cell line(s) for future use.

The significance of the steepness of the concentration-response curve was unclear from the data. The IC_{50} alone does not address this issue. While IC_{20} and IC_{80} (or at least a dose below and above the IC_{50}) were collected for most of the reference substances, they were not used in the analysis. The slope of the concentration-response curve should be included along with the IC_{50} data as additional information about the concentration-response characteristics.

The Draft BRD should include an explanation as to why 3T3 IC_{50} values for numerous reference substances were orders of magnitude different from those determined in the NHK test method. Was this due to cell-specific cytotoxicity? Or was it a consequence of differences in cell culture medium (i.e., presence or absence of serum)?

Table 5-4 in the Draft BRD was highly confusing. The column labeled "Difference (Orders of Magnitude)" contained the calculated ratios of the 3T3/NHK mean IC_{50} values. However, the column contained several mistakes. For example, potassium cyanide, with IC_{50} values of 34.6 vs. 29.0 μg/mL (ratio=1.2), has a *difference* of *1 order of magnitude* while parathion, 37.4 vs. 30.3 (ratio=1.2), has a *difference* of *0*. There were several more such cases (e.g., phenol, carbamazepine, nicotine). A more useful column to compare materials across the two NRU test methods would show the relative difference from the positive control. Since Table 5-5 uses some of the same data as Table 5-4, it must also be revised.
Noted in the summary but not discussed in Section 5.4 were the results in Table 5-4 showing that the IC_{50} values for aminopterin and digoxin differed by five orders of magnitude when tested in 3T3 versus NHK cells. Aminopterin and digoxin are established substrates for organic anionic transporters (OATs). Such transporters are very important for *in vivo* toxicity responses in terms of the ability of molecules to be absorbed, reach target tissues, accumulate, be excreted or secreted. Expression, induction, interference and binding to OATs can strongly influence the *in vivo* effects of a compound. Single nucleotide polymorphisms, which can strongly affect normal function, have been identified in human OATs. The differential susceptibility of the two studied cell lines could be explained by differential functioning of OATs between the cell types, but that was not examined or discussed. At least one publication indicated that NHK cells have at least five different OAT subclass members, with one shown to bind digoxin but not be constitutively expressed in the NHK, which could explain their sensitivity to this chemical. This issue needs to be addressed.

The summary indicated that the IC_{50} values were commonly (92%) within one order of magnitude of each other. A more descriptive and helpful summary would include the fraction

that was within specific IC_{50} ranges. For example, "for nine substances ratios between 3T3 IC_{50} values and NHK IC_{50} values were ≥10 or 0.1, respectively."

5.5 Coded Reference Substances and GLP Guidelines

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.6 Study Timeline and NICEATM/ECVAM Study Participatory Laboratories

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.7 Availability of Data

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.8 Solubility Test Results

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

5.9 Summary

One approach for comparing data generated on the same substance in different laboratories would be to normalize the data using the relative IC_{50} ratios between the reference substances and the positive control (at the level of the individual laboratory). This approach should be considered.

6.0 Accuracy of the 3T3 and NHK NRU Test Methods

This section adequately summarized the accuracy of the studies. The performance and limitations of the two NRU basal cytotoxicity tests were well defined. The overall accuracy for the prediction of Globally Harmonized System (GHS; UN 2005) acute oral hazard category was modest, and enhancement of accuracy through material selection (modular approach), model refinement, or tiered testing strategy should be pursued. Further performance at the extremes of LD_{50} should be considered. Although some analysis of accuracy was conducted related to physical-chemical properties (e.g., solubility) and absorption, distribution, metabolism, and excretion (ADME) (e.g., biotransformation), and other factors (e.g., surface active properties, protein binding, receptor mediation) should be assessed to refine the test methods or draw greater precision by using a modular approach to define the types of materials suitable for the test methods.

Although there was not a significant difference between rat and mouse LD_{50} data (because of the variability of the data), separation was useful because it decreased the biological variability associated with species differences.

6.1 Accuracy of the 3T3 and NHK NRU Test Methods for Predicting Acute Oral Systemic Toxicity

Graphs should be added to compare the responses of the 58 RC substances to the same substances when tested using the 3T3 and NHK NRU test methods.

6.2 Improving the Prediction of *In Vivo* Rodent LD_{50} Values from *In Vitro* NRU IC_{50} Data

6.2.1 *The RC Rat-Only Regression in Millimolar Units*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.2.2 *The RC Rat-Only Regression in Weight Units*

Optimization of the IC_{50}-LD_{50} regression to allow for testing of mixtures was undertaken, yet no mixtures were used in fitting the regression curve. Since the test methods have limitations in accurately predicting the toxicity of materials with known or uncertain mechanisms, the testing of mixtures seems highly controversial.

6.2.3 *The RC Rat-Only Regression in Weight Units Excluding Substances with Specific Mechanisms of Toxicity*

It is true that many of the reference substances with underpredicted toxicity had mechanisms of toxicity that could not be expected to be detected in the 3T3 and NHK cell cultures; however, the Draft BRD incorrectly identified the mechanisms inactive in the cell cultures. The Draft BRD indicated that neurotoxic and cardiotoxic mechanisms, interference with energy utilization, and alkylation of macromolecules would not be active in the cell cultures. Interference with energy metabolism and alkylation of proteins and deoxyribonucleic acid (DNA) actually represent important mechanisms of cytotoxic action, which, in principle, should be detected by cytotoxicity assays such as the 3T3 and NHK NRU test methods. The rationale for excluding the 50 substances with *specific mechanisms of action* appears very questionable. Indeed, Table 6-2 of the Draft BRD shows that the linear regression between rodent LD_{50} values and IC_{50} values was not improved by the exclusion of these substances (R^2=0.353).

In addition, errors were made in the exclusion process based on the rules cited in the Draft BRD. For example, triethylene melamine and busulfan are both alkylating agents, but were not excluded. Paraquat and potassium cyanide were excluded based on interference with energy utilization. However, arsenic trioxide, which can uncouple oxidative phosphorylation, should have been excluded, but was not. Paraquat and potassium cyanide exert their acute systemic toxicity by means of cytotoxic action and should not have been excluded. If using a modular approach based upon proposed mechanisms (e.g., all substances interfering with energy metabolism), then hexachlorophene (a potent uncoupler of mitochondrial

phosphorylation), digoxin (a cardiac glycoside), or propanolol (a β-blocker) should have been included.

The Panel recommended against excluding reference substances based on mechanism given the numerous mechanisms of induction of cytotoxicity, the poor mechanistic understanding of the acute toxicity of many of these materials, and the incomplete knowledge of the appropriateness of the models for the individual modes/mechanisms of action.

6.3 Accuracy of the 3T3 and NHK NRU Test Methods for Toxicity Category Predictions

There was general consensus that adequate data were generated to draw conclusions about the accuracy and validity of the methods. The statistical approaches adopted to analyze data enable accurate and scientifically robust analyses of the two methods with regards to all their aspects.

The evaluation of the accuracy of the NRU basal cytotoxicity test methods for estimating GHS acute oral toxicity category was very extensive and detailed, and it identified areas of concern relative to specific chemical classes, chemicals with known mechanisms of toxicity and particular properties such as solubility, volatility, and so on. The evaluation of concordance of the observed and predicted GHS toxicity categories for each substance was performed correctly. Although a modular approach for using the model may be more reliable, the database was probably too small for most mechanisms of action to draw sound conclusions regarding strengths and limitations of the test methods with respect to chemical classes, mechanisms of toxicity, or physico-chemical properties. Since a mode of action is unlikely to be known about a random source material, it is also unlikely that a modular approach based upon mechanism will often be a viable option. A better approach would be a modular approach to validation based on chemical class, implying similar mode of action. Thus, the justification for the exclusion of 21 substances with specific modes of toxicity was not appropriate. The 26% accuracy for prediction of GHS class without removal of the 21 substances was poor, but better than a random selection using the 72 chemicals (1/6 accuracy).

Corrosivity was an exclusionary criterion intended to be applied to the selection of reference substances (see Section 3 of the Draft BRD). However, corrosive materials as a class were not subsequently deleted from the data when the regression curves were made. Corrosive chemicals are excluded from testing in *in vivo* acute toxicity tests because testing such chemicals *in vivo* is not appropriate, but using data for such chemicals in these analyses is acceptable.

For those classes of substances found to be appropriate for the assay, the NRU-based test methods may also be useful in a development context. During industry screening of new materials, a tool such as this may be useful to rank compounds belonging to the same chemical class (e.g., early lead optimization phase of drug development).

6.3.1 *Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Millimole Regression*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.3.2 *Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods Using the RC Rat-Only Weight Regression*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.3.3 *Prediction of Toxicity Category by the 3T3 and NHK NRU Test Methods with the RC Rat-Only Weight Regression Excluding Substances with Specific Mechanisms of Toxicity*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.3.4 *Summary of the Regressions Evaluated*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.4 Strengths and Limitations of the *In Vitro* NRU Test Methods for *In Vivo* Toxicity Prediction

Use of metabolically competent systems was recommended as one approach to improve the accuracy of *in vitro* predictions of acute toxicity; this should be explored in the future. The use of metabolizing systems is a general requirement for all *in vitro* tests for the prediction of genetic and carcinogenic potential and is considered necessary and scientifically justified. However, the contribution of metabolism of the reference substances was likely misstated, given the incomplete understanding of the acute toxicity of many of them. The substances listed in Table 3-7 of the Draft BRD, which were noted in the analysis of discordant substances, were highly variable in structure and purported mechanism. Of this set of substances, several (e.g., phthalates, valproic acid) may have active metabolites that contribute to their chronic toxicologic effects but which play little or no role in their acute toxicologic effects. Conversely, one may speculate that there may be substances *not* included in Table 3-7 of the Draft BRD for which active metabolism was an important component of its acute effects. Therefore, a more robust analysis of the contribution of metabolism to the accuracy of the models is recommended by incorporating a metabolic system into the *in vitro* assays.

As a future task, the properties of the cell lines (e.g., metabolism, receptors, transporters) that are important for basal cytotoxicity should be better characterized. Identified important properties could be used as performance standards.

6.5 Salient Issues of Data Interpretation

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

6.6 Comparison to Established Performance Standards

It would be informative to show comparisons of the RC LD_{50} values for the selected reference substances used in this study versus the individual laboratory responses for each test instead of the data shown in Figures 6-6 to 6-8 of the Draft BRD, which compares the *in vitro* responses to the overall RC millimole regression data.

While the mean IC_{50} values from one laboratory were generally higher than the rest, comparison to regressions with animal data (Appendix J) suggest there are no major differences between the laboratories in their ability to predict LD_{50} values. In fact the responses in Figures 6-6 to 6-8 look similar. When the *in vitro* response data from all laboratories with the agents selected from the RC are compared to the same agents for the RC, they provide a better correlation with the LD_{50} than did the overall RC data. Given this observation coupled with the variability in the data from animal studies, the data from the *in vitro* test methods would suggest that, as long as the appropriate controls (VC and PC) are used, the data from valid assays should be fairly predictive of animal response. It would be informative to show comparisons of the regression lines using the RC data for the 11 agents shown versus the individual laboratory responses for each test method instead of the data shown in Figures 6-6 to 6-8, which compares the *in vitro* responses to the overall RC millimole regression.

6.7 Summary

Protein binding should be taken into account in the data analyses. This parameter could be eventually taken into account in an additional data analysis (i.e., to the extent possible, consider the free fraction in serum corresponding to the LD_{50} dose). The Hill function slope data and LD_{50} slope data should be compared.

7.0 Reliability of the 3T3 and NHK NRU Test Methods

In general, the analyses in Section 7 adequately addressed the issues regarding both intra- and inter-laboratory reproducibility for the 3T3 and NHK NRU test methods. It was a little bit surprising, however, that some laboratories failed to obtain IC_{50} results for some of the reference substances. The Draft BRD should include an explanation or at least a discussion of these discrepancies, which may relate to the solvent protocol (discussed later). The compounds failing to yield IC_{50} values were mostly solvents (carbon tetrachloride, methanol, xylene, and 1,1,1-trichloroethane). Solvents are an important class of industrial substances for which Toxic Substances Control Act (TSCA) applies. The Draft BRD should offer an explanation if possible. Additional IC_{50} data are available for three of these substances: methanol (1000 mM), 1,1,1-trichloroethane (5.6 mM), and carbon tetrachloride (4.8 mM) using 3T3 cells after 24 hours of exposure (Gülden et al. 2005).

1229
1230 7.1 Substances Used to Determine the Reliability of the 3T3 and NHK NRU Test
1231 Methods
1232
1233 The Panel found that the Draft BRD discussion and evaluation in this section was
1234 appropriate.
1235
1236 7.2 Reproducibility Analyses for the 3T3 and NHK NRU Test Methods
1237
1238 Additional consideration as to the underlying reasons for the variability between the
1239 laboratories would be helpful. The issue of intra- and inter-laboratory reproducibility due to
1240 variations in laboratory practices was addressed during the study and the findings indicated
1241 that the data from the two laboratories with GLP compliant procedures were in closer
1242 agreement and tended to show less variability and lower error rates than the other laboratory
1243 (which had an error rate of 93% for Phases 1a and 1b). Following a common training session
1244 for all laboratories, the interlaboratory variability decreased. This indicates the need for
1245 training in basic methodology and emphasis on protocol compliance. Everyone participating
1246 in such studies should be adequately trained in the basics of cell and tissue culture and sound
1247 scientific methods.
1248
1249 In order to increase the transparency of the comparison of the results from the different
1250 laboratories, an additional analysis of the IC_{50} data could be added: for each substance and
1251 NRU test method, the ratio between the highest and the lowest mean IC_{50} from the
1252 laboratories should be calculated. Those reference substances having ratios ≥ 3.0 should be
1253 presented in a separate table together with their calculated ratios and the names of the
1254 laboratories that delivered the corresponding IC_{50} values. From the Panel's analysis, it
1255 appeared that 17 substances for the 3T3 NRU test method and 11 substances for the NHK
1256 NRU test method had ratios ≥ 3.0. Extreme cases were cupric sulfate with a ratio of 22 (3T3
1257 NRU test method) and digoxin with a ratio of 107 (NHK NRU test method). Furthermore, it
1258 became apparent that even for a simple compound such as sodium chloride, the results from
1259 different laboratories deviated by a factor of more than 3.0 for the NHK NRU test method.
1260
1261 It would be helpful to include a figure in the Draft BRD depicting all IC_{50} values for each test
1262 substance from all laboratories. Graphing of IC_{50} values plus-or-minus (±) the standard
1263 deviation (SD) and rat LD_{50} values ± SD should provide a better comparison of variation in
1264 the two sets of values.
1265
1266 It might also be helpful to look at ratios of the maximum IC_{50} values to the minimum IC_{50}
1267 values to see how they compare *vs.* rodent LD_{50} values. Given the variability in animal data
1268 where LD_{50} values (when more than one LD_{50} was available) could differ from 4 to 14 fold,
1269 the determination of a *precise* IC_{50} in each of the test methods to facilitate the selection of a
1270 starting dose does not seem necessary. Although the comparison of intra- and interlaboratory
1271 reproducibility for the purpose of validating the initial performance was appropriate, the use
1272 of multiple, costly test methods to identify *precise* IC_{50} values to establish initial doses for
1273 determining LD_{50} values seems counterproductive on the basis of cost and would limit
1274 acceptance of such methods.

For some of the reference substances, there was only one point and possibly even no points between 0 and 100% viability. These substances should be identified in the BRD.

NHK NRU IC_{50} data had a better correlation with human LC_{50} values (R^2=0.62) than did rodent 3T3 NRU IC_{50} data (R^2=0.51), as reported by Casati et al. (2005) at the 5th World Congress in Berlin in 2005. The correlation of NHK NRU IC_{50} data with human LC_{50} values (R^2=0.62) was also better than the correlation of rodent LD_{50} data with human LC_{50} values (R^2=0.56) (Casati et al. 2005). Discussion of this relationship should be considered for inclusion in the BRD.

7.2.1 *ANOVA Results for the 3T3 and NHK NRU Test Methods*

The Panel questioned the utility of the ANOVA for addressing the issue of intra- and inter-laboratory reproducibility. Depending upon the sample size and intralaboratory variation, a significant difference could correspond to a very small variation between laboratories or a non-significant difference could correspond to a very large difference between laboratories. Examples include parathion and procainamide. Parathion had reported IC_{50} values of 22.7, 141, and 22 µg/mL (p=0.014, not significant), and procainamide had reported IC_{50} values of 400, 431, and 497 µg/mL (p=0.007, significant). As a consequence, procainamide with satisfying, low interlaboratory reproducibility was included in Table 7-4 (because the ANOVA indicated significant laboratory differences) while parathion was not. There were more such examples that make the utility of the ANOVA questionable.

Based on the ANOVA analysis performed, FAL reported significantly different results from the two other laboratories for 20 substances (3T3 NRU test method). For 18 of these substances FAL reported the highest values. This phenomenon should be explained.

The statistically significant differences among the laboratories for 26 of the reference substances in the 3T3 NRU was worth noting, especially since it was greater than 1/3 of the agents tested. Volatility and/or presence of a precipitate were only noted for nine agents.

7.2.2 *CV Results for the 3T3 and NHK NRU Test Methods*

This section adequately elucidated associations between intra- or interlaboratory reproducibility and chemical classes, chemical properties, and potency categories. The result was that there were no clear associations between any of these parameters and CV values. What was evident, however, was that the reproducibility of both methods depends on the laboratory performing the measurements. A discussion of the possible reasons for this laboratory-specific reproducibility would be helpful.

7.2.3 *Comparison of Laboratory-Specific Linear Regression Analyses for the Prediction of In Vivo Rodent LD_{50} Values from In Vitro NRU IC_{50} Values*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

7.2.4 *Laboratory Concordance for the Prediction of GHS Acute Oral Toxicity Category*
The most important information given here was how often the data generated by the different laboratories would produce different starting doses for the ATC or UDP.

7.3 Historical Positive Control Data

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

7.4 Laboratory Concordance for Solvent Selection

Concern was raised about the differences in solvent selection between laboratories as compared to the BioReliance solvent information. For whatever reason, the variability between laboratories in the selection of solvent pointed out a possible flaw in the solvent determination protocol. This should be evaluated for future studies.

7.5 Summary

Irrespective of the statistical method used (ANOVA or calculation of the ratio between maximum and minimum IC_{50}), there were many more reference substances with deviating results between laboratories in the 3T3 NRU test method than in the NHK NRU test method. This should be explained.

8.0 3T3 and NHK NRU Test Method Data Quality

Section 8 adequately addressed the purpose of this section. No additional data are needed.

8.1 Adherence to Good Laboratory Practice Guidelines

8.1.1 *Guidelines Followed for In Vitro NRU Cytotoxicity Testing*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

8.1.2 *Quality Assurance (QA) for In Vitro NRU Cytotoxicity Test Data*
The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

8.1.3 *Guidelines Followed for In Vivo Rodent Oral LD_{50} Data Collection*
The use of the NRU test relied on the relationship between rat LD_{50} data and the observed IC_{50}. This relationship required reliable LD_{50} measurements for the RC substances used to construct the regression line. Since the LD_{50} values reported by the Registry of Toxic Effects for Chemicals Substances (RTECS®) were the most toxic found in the literature, one is unsure to what extent these LD_{50} estimates can be considered the *gold-standard*. These estimates may be appropriate for risk assessment but these extreme values can be unreliable and could lead to a misleading model of the desired linear relationship.

For comparative purposes with the IC_{50} values, LD_{50} values should reflect the variation observed. In most cases, a range of values should be shown. Such a range should reflect reasonable data with outliers omitted. If no range is shown, then a mean value (when available) plus-or-minus (±) SD should be used for the LD_{50}. The variability in animal data is usually much greater than that found *in vitro*. Therefore, comparing IC_{50} ± SD and Rat LD_{50} ± SD or data range should provide a better comparison. The Panel recommended that these data be shown in the report possibly in a bar graph similar to those in Figure 5-1. Based on the current data, it was not anticipated to have a major effect of the predictive potential of the two *in vitro* test methods. However, it could be important for future studies with other substances. The positive control response limits for a definitive test in Phase III was IC_{50} ± 2.5 SD. If the positive control showed this amount of variation, then why should the reference substances be expected to show any less? The test methods were not designed to predict hazard class but to predict starting animal dose in the acute LD_{50} tests.

8.2 Results of Data Quality Audits

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

8.3 Impact of Deviations from GLPs/Non-compliance

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

8.4 Availability of Laboratory Notebooks

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

8.5 Summary

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

9.0 Other Scientific Reports and Reviews of *In Vitro* Cytotoxicity Test Methods and the Ability of These Test Methods to Predict Acute Systemic Toxicity

In general, reports on other *in vitro* test methods using NRU were useful in providing insights into the correlation as well as the disparities between *in vitro* IC_{50} and *in vivo* LD_{50}. This was particularly true for the previously reported attempts to compare *in vitro* toxicity to *in vivo* lethality. However, it was less clear that the comparisons between eye irritation and NRU *in vitro* test methods were of use in interpreting the data used to compare *in vitro* IC_{50} to *in vivo* LD_{50}. While the mode of exposure is much more comparable between the *in vitro* test methods and the eye irritation (i.e., the test substance is applied directly to the target cell population), the endpoint is dissimilar. Furthermore, direct exposure of the target cells often

cannot adequately predict systemic effects, especially for some classes of substances that act through a known mechanism that does not relate to basal cytotoxicity.

Care was taken in the NICEATM/ECVAM study to cover a range of potencies and mode of action was also considered. It would be useful to compare the range of *in vivo* toxicities and modes of action represented in the other studies reported in Section 9 with the present NICEATM/ECVAM study.

9.1 Relevant Studies

9.1.1 *Correlation of In Vitro NRU Cytotoxicity Results with Rodent Lethality*

Additional discussion from the published literature about the advantages and limitations of using various supplemental metabolizing systems in cell culture for cytotoxicity testing could be included. For the Peloux *et al.* (1992) study, it may be worth including a discussion about the high correlation and whether the relatively good predictive value was a result of the route of exposure (i.e., intravenous [iv] and intraperitoneal [ip]). It should be clarified that the goodness of correlation for the *in vivo/in vitro* values for the different routes of exposure was iv>ip>oral and reflected different kinetic variables.

The results of the workshop presented in Seibert et al. 1996 should be included.

9.1.2 *Use of Cytotoxicity Data to Reduce the Use of Animals in Acute Toxicity Testing*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

9.1.3 *Other Evaluations of 3T3 or NHK NRU Test Methods*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

9.2 Independent Scientific Reviews

9.2.1 *Use of In Vitro Cytotoxicity Data for Estimation of Starting Doses for Acute Oral Toxicity Testing*

Clarification about the percentage reduction of animal use as referenced in the ICCVAM 2001a report should be included in Section 9 with the present ICCVAM study (i.e., what is the likely basis for the difference between then and now).

9.2.2 *Validation of 3T3 NRU for Phototoxicity*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

9.3 Studies Using *In Vitro* Cytotoxicity Test Methods with Established Performance Standards

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

1456
1457 9.4 Summary
1458
1459 The Panel found that the Draft BRD discussion and evaluation in this section was
1460 appropriate.
1461
1462 **10.0 Animal Welfare Considerations (Refinement, Reduction, and Replacement)**
1463
1464 The extent to which the NRU-based methods could contribute to a reduction in animal use
1465 was clearly discussed. The statistical analyses were clearly presented and the conclusions are
1466 clear. However, the Panel indicated that the extent to which the NRU test methods will
1467 reduce animal use for *in vivo* testing was not adequately characterized and discussed. The
1468 calculated savings (8-21%) of animals was only valid if several assumptions were accepted.
1469 For example, 21 of the 72 reference substances were excluded from the calculations because
1470 of their assumed specific modes of action. The best way to evaluate a possible reduction in
1471 animal use by using *in vitro* cytotoxicity to set the starting dose of an unknown substance is
1472 to assume that nothing is known about the mechanism(s) of toxicity of that test chemical.
1473 Therefore, all 72 reference substances should be included in the calculation of animal
1474 savings, regardless of their mode of action.
1475
1476 The use of the NRU cytotoxicity test methods are warranted not only if the number of
1477 animals used in the studies is reduced but also if the stress resulting from chemical exposure
1478 is minimized. The decision to use the NRU test to determine the starting dose for the ATC
1479 method or UDP is justified by the reduction in the number of animals required as indicated in
1480 the simulation studies.
1481
1482 The simulation studies compared the numbers of animals used with the starting dose
1483 indicated by the NRU basal cytotoxicity test method with the numbers of animals used with
1484 the default starting dose. Although the reduction in animals was not that great on a
1485 percentage basis, the testing of 4000 chemicals coming on the market in a year, could save
1486 4000 rats at a rate of one rat per chemical. The Panel indicated, however, that a requirement
1487 to use the NRU test to determine the starting dose could lead to an increase in the number of
1488 animals required particularly if other data were available to provide a more accurate starting
1489 dose.
1490
1491 More information on the doses at which the reductions in expected animal numbers were
1492 found should be provided in the Draft BRD. Presumably, for the most toxic substances, the
1493 savings were at higher doses (as with the NRU test, the starting dose was below the default)
1494 and for the least toxic substances the savings were at the lower doses. The former are more
1495 important than the latter. For the most toxic substances, the largest savings in animal
1496 numbers was provided by the RC millimole regression. This was in contrast to the overall
1497 animal savings, which was smallest when this prediction is used. If the aim was to prevent
1498 animal suffering rather than to reduce animal numbers, then it appeared that the RC
1499 millimole regression was preferable.
1500

10.1 Use of 3T3 and NHK NRU Test Methods to Predict Starting Doses for Acute Systemic Toxicity Assays

This section should clarify that the NRU methods are to be used in a weight-of-evidence approach to determining the starting dose for acute oral systemic toxicity assays. Concern was expressed that underprediction of the toxicity by the cytotoxicity tests might lead to increased animal suffering. Although the accuracy for predicting the exact GHS category appears to be low, the data demonstrates that there is a reduction in animal use versus starting at the default starting dose if no other information is available (e.g., no toxicity information from chemicals with similar structure, etc.).

10.2 Reduction and Refinement of Animal Use for the UDP

Based on existing data, where molecular weight information was available for a relatively pure test substance, the millimolar regression should be used; in the absence of such data, the mg/kg regression should be used.

10.3 Reduction and Refinement of Animal Use for the ATC

The Panel found the discussion and evaluation of this section to be appropriate.

10.4 Summary

The Panel found the discussion and evaluation of this section to be appropriate.

The possibility of using the NRU test methods to determine the starting dose for the fixed dose procedure (FDP) acute toxicity test should be evaluated.

Animal savings should take into account, to the extent possible, the prevalence of chemicals in each GHS category.

11.0 Practical Considerations

Section 11 contained evaluations of potential expense to be incurred upon approval and required implementation of these procedures to aid in choosing the starting dose for a UDP or other type of rat oral toxicity study. However, a cost-benefit analysis was absent. In order to reduce the animal usage per acute oral toxicity study by approximately 1-2 rats, the estimated cost to sponsors increased by $1000-2000 for the preliminary *in vitro* study. This is not cost-effective. Obviously, additional time would be required also to complete the oral toxicity evaluation. Furthermore, although it was said that defining a starting dose to more closely coincide with the actual LD_{50} of a test substance improves the ultimate LD_{50} estimate, many regulatory tests are limit tests for which a preliminary *in vitro* test would offer no benefit.

11.1 Transferability of the 3T3 and NHK NRU Test Methods

It appears that transferability was not as easy as was stated; minor protocol differences can have profound effects. Adequate training must be conducted prior to the initiation of the study, and a demonstration of proficiency in running the test must be demonstrated before testing unknowns.

11.1.1 *Facilities and Major Fixed Equipment*

A dedicated cell culture laboratory should be added to the list of needs.

11.1.2 *Availability of Other Necessary Equipment and Supplies*

A single source for NHK medium was noted to be a problem in the NICEATM/ECVAM validation study.

Although the Draft BRD indicated that laboratories could isolate keratinocytes from donated cultures, this could increase intralaboratory variation. The Panel agreed that the recommendation for a commercial source is better.

The Draft BRD should indicate that it is necessary to confirm that cells are free from contamination (e.g., bacteria, mycoplasma).

11.2 3T3 and NHK NRU Test Method Training Considerations

11.2.1 *Required Training and Expertise*

This section noted that good cell culture practices are needed. The Panel recommended removing statements about the need for training in cloning, transfection, expression cloning, immortalization, and virus propagation since these techniques are not necessary for cytotoxicity testing.

11.2.2 *Training Requirements to Demonstrate Proficiency*

The Panel found that the Draft BRD discussion and evaluation in this section was appropriate.

11.3 Test Method Cost Considerations

11.3.1 *3T3 and NHK NRU Test Methods*

The Panel indicated that the costs quoted may be more than a little bit low. The Draft BRD noted that it was possible that there wouldn't be cost savings using NRU testing first, if only a few rats were used. Additionally, the NHK NRU test could be almost cost-prohibitive if 5 x $380 vials are needed per 5 x 96-well plates.

The costs of performing NRU testing were charges from commercial laboratories. A rough calculation for the cost of equipment and time need to perform each test might help individual laboratories understand the cost and time of performing the test methods.

1590 11.3.2 *In Vivo Rodent Acute Oral Toxicity Testing*
1591 Since the NRU test methods are to be used for reduction of animal use rather than
1592 replacement, it is appropriate to describe the number of animals that might be reduced in this
1593 section.
1594
1595 11.4 Time Considerations for the 3T3 and NHK NRU Test Methods
1596
1597 Since it takes some time to screen the NHK NRU assay medium, it should be described in
1598 this section.
1599
1600 11.5 Summary
1601
1602 The commentaries in Section 11 appeared to be appropriate. It was difficult to compare the
1603 value of the *in vitro* NRU test method ($1120-$1850) per test substance to achieve an IC_{50}
1604 versus an animal test ($750-$3750) to achieve an LD_{50}. If the *in vitro* test can save at least a
1605 single animal in the execution of the ATC or UDP test, this evaluation was worth the effort.
1606

VALIDATION STATUS OF THE NRU TEST METHODS

The Panel agreed that the applicable validation criteria have been adequately addressed for using these *in vitro* test methods in a weight-of-evidence approach to determine the starting dose for acute oral *in vivo* toxicity protocols. However, the Panel was aware that validation of the two NRU test methods was carried out not only to determine if they could be used to set starting doses for *in vivo* acute toxicity studies, but also to determine the extent to which these tests could be a useful step in an *in vitro* tiered testing strategy for acute toxicity. The Panel agreed the validation study showed the two NRU test methods evaluated could not be used as a stand-alone replacement for the *in vivo* tests even considering the variability of the latter. The Panel encouraged future work to develop a tiered testing strategy that includes basal cytotoxicity as part of the overall strategy.

[This Page Intentionally Left Blank]

DRAFT ICCVAM RECOMMENDATIONS FOR *IN VITRO* ACUTE TOXICITY TEST METHODS
(Peer Review Panel Report)

[This Page Intentionally Left Blank]

1.0 Draft ICCVAM Recommendations for *In Vitro* Acute Toxicity Test Methods

1.1 Recommended Test Method Uses

1. The 3T3 and NHK NRU test methods are not sufficiently accurate to predict acute oral toxicity for the purpose of hazard classification (see Section 6 of the *In Vitro* Acute Toxicity Test Methods BRD).

   - The Panel agreed with this statement in that neither of the two basal cytotoxicity tests can be used as alternatives for the *in vivo* acute oral toxicity test for the purposes of hazard classification.
   - In the Draft BRD, the rat *in vivo* data did not conform to current GLP standards.

2. For the purposes of acute oral toxicity testing, the 3T3 and NHK NRU test methods may be used in a weight-of-evidence approach to determine the starting dose for the current acute oral *in vivo* toxicity protocols (i.e., the Up-and-Down Procedure [UDP] and Acute Toxic Class [ATC]).

   - The Panel agreed that the *in vitro* test methods may be useful in a weight-of-evidence approach to determine the starting dose for acute oral *in vivo* toxicity protocols.
   - Given the test methods' limited predictive capacity, however, it was unclear whether they will provide substantial weight in that decision.
   - The overall accuracy was modest, and enhancement of accuracy through material selection (modular approach), model refinement, or tiered testing strategy should be pursued.

3. Consistent with the U.S. Government Principles on the Use of Animals in Research, Testing, and Education (National Research Council 1996), and the U.S. Public Health Service Policy on Humane Care and Use of Laboratory Animals (PHS 2002)[3], *in vitro* basal cytotoxicity test methods as part of a weight-of-evidence approach to estimate the starting dose for acute oral *in vivo* toxicity test methods should be considered and used where appropriate before testing is conducted using animals. For some types of substances, this approach will reduce the number of animals needed. In some testing situations, the approach may also reduce the numbers of animals that die or need to be humanely killed**.**

   - The Panel agreed**.**

[3] National Research Council. 1996. Guide for the Care and Use of Laboratory Animals. Washington, DC: National Academy Press.
PHS. 2002. Public Health Service (PHS) Policy on Humane Care and Use of Laboratory Animals.

4. Substances with specific toxic mechanisms that are not expected to be active in 3T3 or NHK cells (e.g., those that are neurotoxic, cardiotoxic, interfere with energy utilization, or alkylate proteins and other macromolecules) will likely be underpredicted by these *in vitro* basal cytotoxicity test methods. Therefore, until such time as a more predictive testing approach is developed, the results from basal cytotoxicity testing with such substances may not be appropriate.

    - The Panel disagreed with elements of this statement; specific toxic mechanisms that are not expected to be active in 3T3 and NHK cells, such as "interference with energy utilization and alkylation of proteins and other macromolecules", are mechanisms of cytotoxic action and should be detectable with 3T3 and NHK cells.

5. The regression formula used to determine starting doses should be the revised Registry of Cytotoxicity (RC) regression line [with IC_{50} values in μg/mL and LD_{50} values in mg/kg] developed with the RC chemicals using rat LD_{50} data only and excluding chemicals with mechanisms of action that are not expected to be active in *in vitro* basal cytotoxicity test methods.

    - The Panel did not agree with this statement.
    - There was consensus among the Panel that the data contained within the Draft BRD or the open literature were not sufficient to justify the exclusion of reference substances based on mechanism.
    - It was not justified to (retrospectively) exclude substances because of assumed modes of toxic action *in vivo* and/or possible involvement of biotransformation reactions.

6. The performance of other *in vitro* basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD_{50} value, respectively) should be demonstrated to meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods.

    - The Panel agreed with this statement although the reliability of the test methods in this study was not quite satisfying (e.g., inter-laboratory reproducibility), the reproducibility of these methods (e.g., intra-laboratory reproducibility) was modest, and the accuracy of these methods was poor.

7. Compared to the NHK NRU test method, the 3T3 NRU test method appears to be less labor intensive and less expensive to conduct; therefore, the 3T3 NRU cytotoxicity test method is recommended for general use.

    - Some Panel members agreed in a general sense, however, cautioned that one model be preferred over the other, based upon specific knowledge

regarding known mechanisms of action (e.g., the rationale for the disparate results observed with aminopterin and digoxin). Other Panel members agreed with this statement because the use of continuous cell lines is more efficient, especially since the overall animal savings were relatively low.

- One Panel member noted that NHK NRU IC_{50} data have shown a better correlation with human LC_{50} values (R^2=0.62) than do rodent 3T3 NRU IC_{50} data (R^2=0.51) and better than rodent LD_{50} data correlates with human LC_{50} values (R^2=0.56) as reported by Casati et al. (2005) at the 5th World Congress in Berlin. It is important to remember that hazard assessment relates to the safety of humans, not rats.
- Based on costs of commercial keratinocytes, the NHK NRU test method may be cost-prohibitive.
- The proprietary nature of the composition of the NHK culture medium made it impossible to assess the role differences in media composition may have had on the results.

1.2 Draft Recommended Test Method Limitations

- Colored substances (besides red substances) may absorb light in the optical density range of the NRU test methods, which could affect the accuracy of the results.
- The Draft BRD indicated that optimization to allow for testing of mixtures was being undertaken, yet no mixtures were used in fitting the regression curve. Given the limitations of the test methods in accurately predicting materials of known or uncertain mechanisms, the testing of mixtures seems highly controversial.

1.3 Draft Recommended Future Studies

1. Additional data should be collected using the 3T3 and/or the NHK NRU test methods to evaluate their usefulness for predicting the *in vivo* acute oral toxicity of chemical mixtures.

    - The Panel generally agreed that this is a good recommendation, although collecting data could be difficult and doing a correlation with *in vivo* data would be even more difficult. It may be useful to suggest that such data only be collected with the 3T3 NRU test method, and that it would be necessary to clarify the reasons for the interlaboratory variations for future use of the method.

2. Additional high quality comparative *in vitro* basal cytotoxicity data should be collected in tandem with *in vivo* rat acute oral toxicity test results to supplement the high quality validation database started by this study. Periodic evaluations of the expanded database should be conducted to further

characterize the usefulness and limitations of using *in vitro* cytotoxicity data as part of a weight-of-evidence approach to estimate starting doses.

- The Panel agreed this could be valuable under certain conditions, especially if NRU data were collected as acute toxicity testing is conducted.
- However, no reviewer wanted *in vivo* testing conducted solely to collect data to assess the usefulness of the NRU test method, particularly given that the savings in animal numbers that arise from the use of the NRU test method to determine the starting dose for the ATC method or UDP are fairly modest.

3. Additional efforts should be conducted to identify additional *in vitro* tests and other methods necessary to achieve accurate acute oral hazard classification; specifically, studies should be conducted to investigate the potential use of *in vitro* cell-based test methods that incorporate mechanisms of action and evaluations of ADME (absorption, distribution, metabolism, excretion) to provide improved estimates of acute toxicity hazard categories.

- The Panel agreed with this statement and added that there should be additional effort towards development of alternative methods to adequately predict the *in vivo* acute toxicity of chemicals for the purposes of hazard classification.
- An additional statement to include could be, "and the development of methods to extrapolate from *in vitro* toxic concentrations to equivalent doses *in vivo*."

4. The *in vivo* database of reference substances used in this validation study should be used to evaluate the utility of other non-animal approaches to estimate starting doses for acute oral systemic toxicity tests (e.g., widely available software that uses quantitative structure-activity relationships [QSAR]).

- The Panel agreed with this recommendation.

5. Standardized procedures to collect information pertinent to an understanding of the mechanisms of lethality should be included in future *in vivo* rat acute oral toxicity studies. Such information will likely be necessary to support the further development of predictive mechanism-based *in vitro* methods.

- The Panel agreed with this recommendation; this is really important and could further the development of non-animal alternatives in the future.
- To facilitate comparisons and model development, future studies should incorporate high quality animal data for required testing of new substances, blood levels from animals (LC_{50}) (where possible), and high quality *in vitro* data for the same substances.

- To aid in this process, the Panel recommended that an expert group be convened to identify appropriate *in vivo* endpoints.
- The Panel recommended also that ICCVAM consider convening a working group to explore mechanisms of action of acute toxicity, and approaches to acquiring additional information on acute toxic mechanisms when conducting the required *in vivo* acute toxicity testing.
- Although a modular approach may be more reliable, the database was likely too small for most mechanisms of action to draw sound conclusions regarding strengths and limitations of the test methods with respect to chemical classes, mechanisms of toxicity, or physico-chemical properties. Since a mode of action is unlikely to be known about a random source material, it is also unlikely that a modular approach based upon mechanism will be a viable option. A better approach to validation is one based on chemical class, implying similar mode of action.

6. An expanded list of reference substances with estimated rat LD_{50} values substantiated by high quality *in vivo* data should be developed for use in future *in vitro* test method development and validation studies.

- The Panel agreed with this recommendation; there should be a concerted effort to collect proprietary data.

[This Page Intentionally Left Blank]

APPENDIX A
DRAFT PERFORMANCE STANDARDS FOR *IN VITRO* ACUTE TOXICITY METHODS
(Peer Review Panel Report)

[This Page Intentionally Left Blank]

1.0 Purpose and Background of Performance Standards

The available data from this study appeared to support the validity of the recommended performance standards for the test methods. The usefulness and limitations were well covered, and if validated, the methods may be a worthwhile option. However, there may be some cause for concern if use of the methods is made compulsory for regulatory purposes.

1.1 Introduction

The Panel found the discussion and evaluation of this section to be appropriate.

1.2 Elements of ICCVAM Performance Standards

The Panel found the discussion and evaluation of this section to be appropriate.

1.3 ICCVAM Process for the Development of Performance Standards

The Panel found the discussion and evaluation of this section to be appropriate.

1.4 ICCVAM Development of Recommended Performance Standards for *In Vitro* Acute Toxicity Test Methods

The Panel found the discussion and evaluation of this section to be appropriate.

2.0 *In Vitro* Acute Toxicity Test Methods

The Panel found the discussion and evaluation of this section to be appropriate.

2.1 Background

The Panel found the discussion and evaluation of this section to be appropriate.

2.2 Principles of *In Vitro* Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity Tests

The Panel found the discussion and evaluation of this section to be appropriate.

2.3 Essential Test Method Components for *In Vitro* Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity (Lethality) Tests

A discussion is needed in this section regarding whether or not the NRU test methods are recommended for use with unknown substances and mixtures. The recommendations made in Section 2.3.2 (Application of the Test Substances), Section 2.3.3 (Control Substances), and Section 2.3.4 (Viability Measurements) were acceptable.

2.4 Reference Substances for *In Vitro* Basal Cytotoxicity Assays to Predict Starting Doses for Acute Oral Toxicity Tests

The significance of the secondary chemical subset to be used for *investigational purposes* should be better elucidated.

2.5 Accuracy and Reliability

The Panel found the discussion and evaluation of this section to be appropriate.

APPENDIX B
DRAFT RECOMMENDED TEST METHOD PROTOCOLS
(Peer Review Panel Report)

[This Page Intentionally Left Blank]

1.0 Draft Recommended Test Method Protocols

The protocols were generally quite detailed and laboratory technicians should be able follow the procedures. The Panel recommended the following clarifications be added to the 3T3 and NHK NRU test method protocols:

1.1 Protocol Recommendations

- The rationale for testing the positive control on separate plates rather than on the test plates should be provided.
- The number of definitive tests that should be performed for a test substance should be specified.
- The range of linearity of the microplate reader should be confirmed (as per in-house SOPs) for the recommended optical density (OD_{540}) and stated.
- Maximum absorbance values needed by a spectrophotometric plate reader should be provided for application to the NRU test methods.
- The test method protocols should be streamlined. (Undefined is how this should be accomplished.)
- Guidance for using methods other than the Hill function to determine IC_{50} values should be provided.
- The lowest acceptable test substance dilution factor (i.e., 1.21) should be reduced rather than accepting only one cytotoxicity point between 0 and 100% viability on a steep dose-response curve to use for determination of the IC_{50} value.
- Study directors and quality assurance units are necessary only if testing is performed under Good Laboratory Procedures (GLP), which is not usually necessary for dose-setting tests.
- The protocol for the NHK cells should include a statement about the need to avoid allowing the cell to reach confluence: under these conditions, these cells can exhibit contact-induced differentiation. Once differentiation is induced, cells lose their ability to proliferate.

1.2 Cell Culture Recommendations

- Good cell culture practices (e.g., Hartung et al. 2002) must be followed.
- Whether or not a prequalification test of new keratinocytes should be performed by the laboratory prior to actual testing should be stated.
- A recommendation that keratinocytes should be procured only through commercial sources and not by preparing primary cultures from donated tissue should be included.

1.3 Solubility Recommendations

- Additional guidance to the solubility step-wise procedure should be added (i.e., ensure that test substance solution preparation procedures can be easily understood by laboratory technicians).

- Include a recommendation for training laboratory technicians so they better understand solvent and solubility determinations.
- Additional guidance as to the use of a microscope to assist in determining solubility of a test substance should be added.
- Test substances that may etch plastic or *film out* in medium should be identified (the importance of detecting such compounds by the laboratory technicians should be emphasized).
- The protocols should recommend the use of a solvent (e.g., dimethylsulfoxide [DMSO], ethanol) at its lowest possible concentration.
- There was concern about the differences in solvent selection between laboratories as compared to the BioReliance solvent information. The variability between laboratories in the selection of solvent points out a possible flaw in the solvent determination protocol. This should be evaluated for future studies.

REFERENCES

Autian J, Dillingham EO. 1978. Overview of general toxicity testing with emphasis on special tissue culture tests. In: *In Vitro* Toxicity Testing 1975-1976 (Berky J, Sherrod C, eds). Philadelphia:Franklin Institute Press, 23-49.

Casati S, Hoffmann S, Strickland J, Paris M, Stokes W, Tice R, Malerba I, Hartung T. 2005. Anaylsis of the correlation between *in vitro* cytotoxicity data and acute toxic effects in humans. 5th World Congress on Alternatives and Animal Use in the Life Sciences, Berlin, Germany.

Elmore E, Luc TT, Steele VE, Redpath JL. 2001. Comparative tissue-specific toxicities of 20 cancer preventive agents using cultured cells from 8 different normal human epithelia. In Vitr Mol Toxicol 14:191-207.

Elmore E, Siddiqui S, Desai N, Moyer MP, Steele VE, Redpath JL. 2002. The human epithelial cell cytotoxicity test method for determining tissue specific toxicity: method modifications. Methods Cell Sci 24:145-153.

Fentem J, Fry J, Garle M, Gülden M, Seibert H, Voss J-U, Wassermann O, Perchermeier M, Wiebel F. 1993. An international evaluation of selected *in vitro* toxicity test systems for predicting acute systemic toxicity. A report prepared for DGXI, CEC; Contract Numbers B92/B4-3063-4086 & B92/B4-3040/14087. FRAME Nottingham.

Gülden M, Mörchel S, Seibert H. 2005. Comparison of mammalian and fish cell line cytotoxicity: impact of endpoint and exposure duration. Aquat Toxicol 71:229-236.

Hartung T, Balls M, Bardouille C, Blanck O, Coecke S, Gstrauthaler G, Lewis D. 2002. Good Cell Culture Practice: ECVAM Good Cell Culture Practice Task Force Report 1. Altern Lab Anim 30:407-414. Available: http://ecvam.jrc.it/publication/index5007.html. [accessed 30 May 2006].

ICCVAM. 2001a. Report of the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/. [accessed 29 May 2006].

ICCVAM. 2001b. Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity. NIH Publication No. 01-4500. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].

Morrison JK, Quinton RM, Reinert M. 1968. The purpose and value of LD_{50} determination. In: Modern Trends in Toxicology, Vol 1, (Boyland E, Goulding R, eds), Chichester:John Wiley & Sons, 1-17.

Seibert H, Balls M, Fentem JH, Bianchi V, Clothier RH, Dierickx PJ, Ekwall B, Garle MJ, Gómez-Lechón MJ, Gribaldo L, Gülden M, Liebsch M, Rasmussen E, Roguet, Shrivastava R, Walum E. 1996. Acute toxicity testing *in vitro* and the classification and labeling of chemicals: The report and recommendations of ECVAM Workshop 16. Altern Lab Anim 24:499-510.

UN. 2005. Globally Harmonized System of Classification and Labelling of Chemicals (GHS), First Revised Edition. [ST/SG/AC.10/30/Rev.1]. United Nations, New York and Geneva. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev01/01files_e.html [accessed 12 September 2005].

APPENDIX A2

MINUTES FROM PEER REVIEW PANEL MEETING ON MAY 23, 2006

[This Page Intentionally Left Blank]

Meeting Summary
Peer Review Panel Public Meeting
***In Vitro* Methods for Estimating Starting Doses for Acute Systemic Toxicity Testing**
National Institutes of Health (NIH), Natcher Conference Center
Bethesda, MD

May 23, 2006
8:30 a.m. – 5:00 p.m.

Call to Order

Dr. Katherine Stitzel (Panel Chair) called the meeting to order at 8:30 a.m. and asked all Peer Panel members, National Toxicology Program Interagency Center for the Evaluation of Acute Toxicological Methods (NICEATM) staff, members of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the ICCVAM Acute Toxicity Working Group (ATWG) in attendance, and members of the public to state their name and affiliation for the record. She requested that all individuals identify themselves when they spoke and to use the provided microphones. She stated that two public comment periods would be held during the meeting and asked that individuals who wanted to speak, other than those who had pre-registered, to register at the registration table.

Welcome from the Director, NICEATM and Conflict of Interest Statements

Dr. Stitzel introduced Dr. William Stokes, the director of NICEATM. On behalf of the NIEHS and NICEATM, Dr. Stokes welcomed everyone and thanked the participants for agreeing to serve on the Panel. Dr. Stokes stated that he would serve as the Designated Federal Official for the public meeting. He stated that the meeting was being held in accordance with the Federal Advisory Committee Act (FACA) regulations and that the Panel is constituted under the NIH Special Emphasis Panel charter. Dr. Stokes read the conflict of interest statement and asked the Panel members to indicate if they had any conflicts and to recuse themselves from discussion and voting on any aspect of the meeting for they had any conflict. Dr. Daniel Wilson of the Dow Chemical Company stated that his company produces a number of chemicals used in the validation study, but that he did not consider this to constitute a conflict of interest.

Welcome from the ICCVAM Chair

Dr. Leonard Schechtman, U.S. Food and Drug Administration, Chairman of ICCVAM, welcomed everyone on behalf of ICCVAM. He expressed his appreciation for the Panel's willingness to participate in the peer review process and requested input from the Panel on *in vitro* methods for use in estimating the starting dose for acute toxicity testing. He thanked NICEATM staff and the ATWG, and other ICCVAM members for their efforts in developing the materials and draft recommendations being considered at this peer review meeting. He said that the Panel's report will used by ICCVAM in finalizing its recommendations.

Overview of the ICCVAM Test Method Evaluation Process and Charge to the Panel

Dr. Stitzel asked Dr. Stokes to provide an overview of the ICCVAM test method evaluation process. He stated that the international Panel was made up of 16 scientists from six countries

(United States, United Kingdom, Canada, Japan, Germany, and Italy). He described the 15 ICCVAM agencies and reviewed ICCVAM's history and development. Dr. Stokes summarized the preamble of the ICCVAM Authorization Act and detailed the purpose and duties of ICCVAM as prescribed by the Act. He noted that one of ICCVAM's duties is to review and evaluate new, revised and alternative test methods applicable to regulatory testing. Dr. Stokes described the role of NICEATM in conducting validation studies when funds are available. He stated that all of the reports produced by NICEATM are available from the ICCVAM-NICEATM website or directly from NICEATM.

Dr. Stokes stated that validation is performed to determine the usefulness and limitations of a test method for a specific purpose. He continued by stating that validation is defined by ICCVAM as the process by which the *reliability* and *relevance* of a procedure are established for a specific purpose and that adequate validation is a prerequisite for Federal regulatory acceptance. He listed the ICCVAM criteria for test method validation and acceptance. Dr. Stokes explained that acute toxicity testing was necessary to evaluate and classify the hazard potential of acute single exposures to substances. He stated that poisoning is the second leading cause of injury-related death in the United States.

Dr. Stokes briefly reviewed the ICCVAM International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity, which was held in October 2000. The overall goal of the Workshop was to review the then current status of using *in vitro* testing for predicting acute oral toxicity. The workshop recommended that a near-term goal should be to reduce animal use for acute systemic toxicity assays by using *in vitro* methods to estimate starting doses. A long term goal should be to replace animal use with *in vitro* methods that can predict human acute systemic toxicity using human cells and tissues. In addition to a Workshop Report (ICCVAM. 2001. Report Of The International Workshop On In Vitro Methods For Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/methods/invitro.htm.) A Guidance Document was also published (ICCVAM. 2001. Guidance Document On Using In Vitro Data To Estimate In Vivo Starting Doses For Acute Toxicity. NIH Publication No. 01-4500. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/methods/invitro.htm.). This document also provided two standardized *in vitro* basal cytotoxicity protocols that were the basis for those used in the NICEATM/European Centre for the Validation of Alternative Methods (ECVAM) validation study. As a result of the workshop, ICCVAM made recommended that additional research and development should be conducted to develop the *in vitro* systems, in addition to basal cytotoxicity, that will be necessary to accurately predict acute toxicity without animals (e.g., those that can predict absorption, distribution, metabolism, excretion [ADME] and target organ toxicity). The ECVAM-sponsored A-Cute-Tox project is currently working to develop these *in vitro* test systems that will be necessary to develop this strategy.

Charge to the Panel

Dr. Stokes presented the timeline for conduct of the NICEATM/ECVAM validation study and he then reviewed the charge to the Panel: 1) review the BRD for omissions and errors; 2) evaluate the extent to which each of the applicable criteria for validation and acceptance

have been adequately addressed for the test methods and their specific proposed use; and 3) comment on the extent to which the draft ICCVAM test method recommendations are supported by the information provided in the BRD.

Dr. Stokes presented the rosters for the Peer Panel, ICCVAM, ATWG, and NICEATM and acknowledged the three laboratories that participated in the study: 1) U.S. Army Edgewood Chemical Biological Center, 2) Fund for the Replacement of Animals in Medical Experiments [FRAME] Alternatives Laboratory [FAL] and 3) Institute for *In Vitro* Sciences.

Overview of Acute Oral Toxicity Regulatory Testing Requirements, Hazard Classification Schemes, and the Current Acute Oral Toxicity Regulatory Testing Procedures

Dr. Amy Rispin presented the U.S. statutes and regulations requiring acute oral toxicity testing. She emphasized the use of the three Organization for Economic Cooperation and Development (OECD) Acute Oral Toxicity Test Guidelines (TG 425, TG 423, TG 420) that can be used to meet these test requirements. She stated that acute toxicity has been one of the longest standing areas of regulation in the United States and Europe. Regulatory applications include classification and labeling, risk assessment (key area emphasized by the U.S. Consumer Product Safety Commission [CPSC]), and risk management. Applications of acute toxicity testing have driven obligatory use of protective clothing and other improvements in safety with respect to potential chemical exposures. She stated that the United States is in an active transition period along with the rest of the world toward using the United Nations (UN) Globally Harmonized System (GHS) of Classification and Labelling of Chemicals for product labeling. Dr. Rispin described the current hazard classification systems of various regulatory authorities (i.e., U.S. Environmental Protection Agency [EPA], European Union [EU], U.S. CPSC, U.S. Department of Transportation [DOT], UN GHS).

With regard to test methods for acute toxicity testing, Dr. Rispin provided descriptions of the Up-and-Down Procedure (UDP) Limit test, the UDP Main test, the Acute Toxic Class (ATC) method, and the Fixed Dose Procedure (FDP). Dr. Rispin stated that the UDP has the greatest versatility and is the most commonly used method in the United States. The test uses the most sensitive gender of rat. She explained that the default dosing scheme for this method tends to yield a value lower than median LD_{50} value (i.e., the dose of a test substance that produces death in 50% of the animals tested), which provides the most conservative outcome with dosing of fewer animals. Each test method works better with a starting dose near the LD_{50} value. Background information on the test chemical is very helpful to determine the most appropriate starting dose for acute oral toxicity testing but a default starting dose is available for all methods if no other information is available.

Test Method Overview

Dr. Judy Strickland provided an overview of the NICEATM/ECVAM validation study. The study objectives were:

- Further standardize and optimize two *in vitro* neutral red uptake (NRU) cytotoxicity protocols to maximize intra- and inter-laboratory reproducibility
- Estimate the reduction and refinement in animal use from using *in vitro* basal cytotoxicity assays to identify starting doses for *in vivo* acute toxicity tests

- Assess the accuracy of the two standardized *in vitro* cytotoxicity test methods for estimating rodent oral LD_{50} values across the GHS categories of acute oral toxicity
- Generate high quality *in vivo* lethality and *in vitro* cytotoxicity databases that can be used to support the investigation of the other *in vitro* test methods necessary to accurately predict acute systemic toxicity

Dr. Strickland presented the prioritization criteria used for selection of the reference substances used in the validation study (e.g., substances needed human toxicity/exposure data, rodent toxicity data, and should be relatively nonvolatile). She then described the sequence of events involved in the testing of the reference substances. The reference substances were first tested using a solubility protocol and then tested in the *in vitro* NRU assays. She explained the test acceptance criteria used for ascertaining which tests were functioning optimally. A graphical presentation of an *in vitro* NRU dose-response curve was provided to illustrate how the IC_{50} values (i.e., the concentration of a test substance that reduces cell viability by 50%) were calculated. The IC_{50} values were then used in a linear regression equation to predict corresponding LD_{50} values and to estimate the starting doses for the UDP or ATC methods. Dr. Strickland explained that computer simulation modeling of *in vivo* testing was used to determine animal use with either the default starting dose or the NRU-based starting dose. She provided an example for the UDP method. She stated that testing chemicals with an $LD_{50} > 300$ mg/kg and using the NRU-based starting dose would save 1 – 2 animals per test, or about 11 to 20%.

Dr. Strickland acknowledged the members of the Study Management Team, the laboratories and study directors involved in the study, and other support personnel who assisted in the study.

PEER REVIEW PANEL EVALUATION:
(1) Background Review Document (BRD) for Completeness, Errors, and Omissions
(2) Validation Status of the Proposed Test Methods

Dr. Stitzel provided the following statement to the Panel prior to discussions of the BRD: "To ensure adherence to the Federal Peer Review requirements, the Panel is asked to determine the completeness of the BRD and identify any errors or omissions. Additionally, the Panel will: 1) evaluate the validation status of the proposed test methods, and 2) make a determination of whether the information provided in the BRD supports the draft ICCVAM recommendations."

Dr. Stitzel also stated that before the Panel finalized its conclusions and recommendations, there would be an opportunity for public comment. She introduced the relevant Panel Group Leaders for each BRD Section: (Dr. Marion Ehrich - Section 1, 2, and 11; Dr. Daniel Marsman - Section 3, 5, and 6; Dr. Eugene Elmore - Sections 7 and 8; Dr. Andrew Rowan - Sections 4, 9, and 10). The Group Leaders presented the draft responses to the Evaluation Guidance Questions for consideration by the entire Panel.

Proposed Panel Recommendations for the BRD

BRD Section 1
Introduction and Rationale for the Use of *In Vitro* Neutral Red Uptake Cytotoxicity Test Methods to Predict Starting Doses for *In Vivo* Acute Oral Systemic Toxicity Testing

Dr. Ehrich provided a brief summary of Section 1 and listed the group's draft recommended revisions to this section of the BRD.

- The major conclusions from the workshop presented in Seibert et al. 1996 (Acute Toxicity Testing In Vitro and the Classification and Labelling of Chemicals. The Report and Recommendations of ECVAM Workshop 16. Alternatives to Laboratory Animals 24:499-510) should be included.
- The possibility of using the NRU assays to determine the starting doses for the FDP acute toxicity test should be included.
- A better explanation of why the 3T3 and NHK cells were chosen for the study should be provided.
- The 3T3 and NHK cell doubling times should be included (as a range).

Dr. Stitzel asked for comments from the Panel on this section of the BRD. Since no comments were provided, the Panel agreed upon the draft recommended revisions.

BRD Section 2
Test Method Protocol Components of the 3T3 and NHK *In Vitro* NRU Test Methods

Dr. Ehrich provided a brief summary of Section 2 and listed the groups draft recommended revisions to this section of the BRD.

- The rationale for not using heat-inactivated serum in the cell cultures should be presented.
- The rationale for not using 3T3 cells after approximately 18 passages in culture should be provided.
- The extent to which using different lots of NHK cells in different studies may affect test method variability should be discussed.
- The potential for NHK cells under confluence to differentiate should be discussed as this may affect their sensitivity to cytotoxic agents.
- The variability in the composition of the bovine pituitary extract added to the NHK culture medium should be discussed.
- The procedures for preparation of test chemical dilutions should be clarified.
- The extent to whether cells recover and/or divide should be discussed.
- The vehicle control NRU optical density at 540 nm (OD_{540}) ranges for each laboratory should be presented.
- A discussion should be provided as to whether something other than mechanism of action could have contributed to the unusual concentration-response curves.
- The reference substances that used the study's lowest acceptable test chemical dilution factor (i.e., 1.21) should be listed.

- Additional explanations as to how GraphPad Prism® software calculated the IC_{50} using the Hill function should be provided.
- Quantitative data and the extent of variability on the doubling times of the two cell types for all laboratories during initial cell seeding, after seeding the cells in 96 well plates, and during exposure should be included.

Dr. Stitzel asked for discussion and any other revisions from the Panel on this section of the BRD. No further revisions were proposed and the Panel agreed with the draft recommended revisions.

BRD Section 3
Reference Substances Used for Validation of the 3T3 and NHK NRU Test Methods

Dr. Marsman discussed Section 3. He was satisfied with the selection of the reference substances but questioned the selective removal of some reference substances (based on mechanism of action) from the analyses since there was an incomplete understanding of the mechanisms of action for all of the reference substances. He provided additional recommendations for this section and then Dr. Stitzel asked for comments from the Panel.

Dr. Ehrich asked if the outcome would change if more chemical classes were added. Dr. Marsman said that there was an adequate number of chemical classes tested. Dr. Hasso Seibert stated that characterization of the chemicals is important; however, it was not necessary to do a metabolic profile of each chemical in order to do testing but the information would be useful. Dr. Stokes said that it would be valuable to know if there is a standardized approach to getting such information and requested suggestions from the Panel. Dr. Seibert stated that he was unaware of any standardized methods. Dr. Elmore suggested adding octanol:water coefficients for the test substances if known.

Other recommended revisions to this section of the BRD included:

- The basis for the selection of reference substances appears to be well described and of generally high quality. A wide range of substances, belonging to many chemical classes, physical properties, and different types of toxicities have been included. However, there were no polycyclic aromatic hydrocarbons, catalysts, simple aldehydes, ketones, biocides, cosmetic ingredients, plant toxins or other natural compounds. Also, it would have been useful to include some mixtures similar to likely pesticide or household product formulations.
- The adequacy of the range of reference substances and their mechanisms of oral toxicity is difficult to judge as there is often very limited knowledge about their mechanisms of action. Specifically, there is little information about the reference substances to support that specific modes of action of acute systemic toxicity have been robustly explored.
- The molecular structure of each reference substance should be provided.
- The cytotoxicity endpoint for the assay is based on uptake of neutral red into lysosomes; no mention is made whether any of the references substances cause lysosomal swelling, which could cause artifacts. Within this context,

there may be some limited value in adding data from additional substances to improve precision, and inclusion of substances at the extremes of the GHS toxicity categories may be helpful.

- There is some concern that the potential for bias may exist if the reference substances were pre-selected based on best fit to a regression line plotting cytotoxicity versus *in vivo* LD_{50} to evaluate *in vitro* test methods to estimate the acute oral LD_{50}.
- To the extent possible, characterization of the metabolic profiles of the reference substances should be added.
- Several confounding factors have not been addressed in the selection or evaluation of materials. For example, the octanol:water coefficients and the surface-active potential (to the extent possible) should be characterized and this information incorporated into the assessment.

Dr. Stitzel asked for comments from the Panel on this section of the BRD. Since no comments were provided, the Panel agreed upon the draft recommended revisions.

BRD Section 4
***In Vivo* Rodent Toxicity Reference Values Used to Assess the Accuracy of the 3T3 and NHK NRU Test Methods**

Dr. Rowan led the discussion on Section 4 and presented the following recommended revisions to this section of the BRD.

- In general, the *in vivo* acute oral toxicity data did not conform to modern standards of toxicity testing and hence their quality would be difficult to determine.
- The LD_{50} values from the Registry of Toxic Effects for Chemical Substances (RTECS®) used in the Registry of Cytotoxicity (RC) linear regression model may not be the "gold standard" values. Extreme values may be unreliable and could lead to a misleading model of the desired linear relationship.

Dr. Stitzel asked for comments from the Panel on this section of the BRD. Since no comments were provided, the Panel agreed upon the draft recommended revisions.

BRD Section 5
3T3 and NHK NRU Test Method Data and Results

Dr. Marsman presented the recommended revisions to the Panel and then Dr. Stitzel asked for comments from the Panel.

The Panel suggested performing a comparison of cell types, with respect to sensitivity to the individual chemicals, by normalizing the IC_{50} values to the IC_{50} of the positive control (PC). The comparative response of each cell type might elucidate whether an individual chemical is an outlier (with respect to prediction of GHS classification). The concordance of IC_{50} values for the two test methods is basically good since only 3% of the reference substances differed by two orders of magnitude and 3% of the reference substances differed by greater

than five orders of magnitude. It is important to know how these cell types respond to the different chemical classes. This relates to the precision of the test in relation to the GHS classification. A 10-fold difference in IC_{50} values between 3T3 and NHK cells may not pose a problem since biology is not exact, but it is important to know the biological differences since this will help in understanding how the systems work.

Other recommended revisions to the BRD included:

- Explanations, if available, should be added as to why carbon tetrachloride and a few other reference substances could not be adequately tested by all laboratories.
- An explanation, if available, for the considerably higher sensitivity of the NHK versus 3T3 cells to the positive control should be provided.
- Further discussion is needed exploring the biological significance of and possible reasons for the differences in sensitivity and selectivity between the two cell lines; this may be useful for selecting the appropriate cell line(s).
- A descriptive summary of the IC_{50} values and orders of magnitude that includes the fraction that were within specific IC_{50} ranges should be provided.
- The Hill function slope data and LD_{50} slope data should be provided for potential comparisons of IC_{50} slopes to LD_{50} slopes.
- A discussion about why the IC_{50} values for aminopterin and digoxin differ by five orders of magnitude when comparing 3T3 and NHK values should be provided. Information about organic anionic transporters should be included.
- The relative IC_{50} ratios between the reference substances and the positive control (at the level of the individual lab) should be used to compare materials across assays.

Dr. Stitzel asked for comments from the Panel on this section of the BRD. Since no comments were provided, the Panel agreed upon the draft recommended revisions.

BRD Section 6
Accuracy of the 3T3 and NHK NRU Test Methods

Dr. Marsman led the discussion of Section 6.

The Panel was not sure if it is important to separate rat and mouse LD_{50} data but recommends separation because it is more scientifically acceptable. The animal data already has much variability (e.g., age, gender, etc.) and additional variability such as combining data from different species should be avoided.

The Panel addressed the use of corrosive chemicals in the study. A caveat should be added to the BRD that *in vivo* testing of corrosives is neither advocated nor recommended. If, however, historical *in vivo* data on such chemicals exist, the data should used and analyzed in conjunction with *in vitro* data.

There was a consensus that adequate data were generated to draw conclusions about the accuracy and reproducibility of the two test methods. The statistical approaches adopted to

analyze the data enabled accurate and scientifically robust analyses of test method accuracy. The information presented in this section of BRD appears sufficient with the following exceptions, which the Panel recommended as revisions to this section of the BRD:

- The overall accuracy is modest, and enhancement of accuracy through material selection (modular approach), model refinement, or tiered testing strategy should be pursued.
- The basis for the orders of magnitude difference in IC_{50} values for numerous reference substances between 3T3 and NHK cells should be explained (i.e., is the difference a consequence of cell-specific cytotoxicity or differences in exposure conditions or something else?).
- Chemicals in the RC database that showed underprediction of toxicity were deemed to have mechanisms of toxicity that could not be detected in the 3T3 and NHK NRU assays. These mechanisms included neurotoxic and cardiotoxic mechanisms, interference with energy utilization, and alkylation of macromolecules. The Panel indicated that interference with energy metabolism and alkylation of proteins and DNA represent important mechanisms of cytotoxic action. Thus, the rationale for excluding the substances from the RC database with "specific mechanisms of action" appears very questionable (i.e., all chemicals should remain in the regression).
- Given that a mode of action is unlikely to be known about a random material, a modular approach based upon mechanism is not a viable option. A better approach would be one based on chemical class, implying similar mode of action.
- Use of metabolically competent systems is recommended as one approach to improve the accuracy of *in vitro* predictions of acute toxicity and should be explored in the future.
- Corrosivity was one of the exclusionary criteria that was originally attempted to be applied to the reference substances. However, corrosive materials as a class were not deleted from calculation of the regression lines.
- Graphs should be added to compare the responses of the 58 RC substances to the same agents with the 3T3 and NHK NRU tests.
- The criterion for removal of some substances to arrive at the best regression is of limited merit; however, without removal, the 26% accuracy for prediction of GHS class is poor although better than a random selection using the 72 chemicals (1/6 accuracy).
- As a future task, the properties of the cell lines (e.g., metabolism, receptors, transporters) that are important for basal cytotoxicity should be better characterized. These properties could be used in performance standards.
- The proprietary nature of the composition of the NHK culture medium makes it impossible to assess the role differences in media composition may have had on the results.
- It would be informative to show comparisons of the regressions (using RC IC_{50} and LD_{50} data) for the selected agents used in this study versus the individual lab responses for each test instead of the data shown in Figures 6-6 to 6-8 of the BRD, which compares the *in vitro* responses to the overall RC millimolar (mM) regression.

- Protein binding should be taken into account in additional analyses (i.e., to the extent possible, consider the free fraction in serum that corresponds to the LD_{50} dose).
- The Hill function slope data and LD_{50} slope data should be compared.
- Graphing of IC_{50} values ± the standard deviation (SD) and rat LD_{50} values ± SD should provide a better comparison of variation in the two sets of values.

Dr. Stitzel asked for comments from the Panel on this section of the BRD. Since no comments were provided, the Panel agreed upon the draft recommended revisions.

BRD Section 7
Reliability of the 3T3 and NHK NRU Test Methods

Dr. Elmore led the discussion of Section 7 in regard to the draft recommended revisions to this section of the BRD.

- Additional consideration as to the underlying reasons for the variability between the laboratories would be helpful. Everyone participating in these studies should be adequately trained in the basics of cell and tissue culture and sound scientific methods.
- This section adequately elucidated associations between intra- or inter-laboratory reproducibility and chemical classes, chemical properties, and potency categories; there were no clear associations between any of these parameters and coefficient of variation (CV values). However, the reproducibility of both methods depends on the laboratory performing the measurements. A discussion of the possible reasons for this laboratory-specific reproducibility would be helpful.
- IC_{50} values do not indicate the steepness of the concentration-response curve. IC_{20} (i.e., the concentration of a test substances that reduces cell viability by 20%) and IC_{80} (i.e., the concentration of a test substances that reduces cell viability by 80%) values were collected, but not used. For some reference substances, there was only one point between 0 and 100% viability.
- The reference substances failing to yield IC_{50} values were mostly solvents (e.g., carbon tetrachloride, methanol, xylene, trichloroethane). An explanation should be provided.
- The Panel questioned the utility of the analysis of variance analysis (ANOVA) for addressing the issue of intra- and inter-laboratory reproducibility. Depending upon the sample size and intralaboratory variation, a significant difference could correspond to a very small variation between laboratories or a nonsignificant difference could correspond to a very large difference between laboratories. The content of Table 7-4 should be examined to assure that the correct data are included.
- Based on the ANOVA analysis performed, FAL reported significantly different results from the two other laboratories for 20 substances (3T3 NRU assay) and for 18 of these substances FAL reported the highest values. The BRD should explain this phenomenon.

- Independent of the statistical method used, there were more reference substances with deviating results between laboratories for the 3T3 NRU assay than for the NHK NRU assay. The BRD should explain this.
- The BRD should explain why some laboratories failed to obtain IC_{50} results for some reference substances.
- It might be helpful to look at ratios of the maximum IC_{50} values to minimum IC_{50} values to see how they compare vs. rodent LD_{50} values. Those chemicals having ratios ≥ 3.0 should be presented in a separate table together with their calculated ratios and the names of the labs that delivered the corresponding IC_{50} values.

Dr. Stitzel asked for comments from the Panel on this section of the BRD. Since no comments were provided, the Panel agreed upon the draft recommended revisions.

BRD Section 8
3T3 and NHK NRU Test Method Data Quality

Dr. Elmore led the discussion of Section 8. The Panel did not recommend any revisions to this section of the BRD. Dr. Stitzel asked for comments from the Panel; the Panel accepted the draft decision to not recommend revisions to Section 8.

BRD Section 9
Other Scientific Reports and Reviews of *In Vitro* Cytotoxicity Test Methods and the Ability of These Test Methods to Predict Acute Systemic Toxicity

Dr. Rowan led the discussion of Section 9 on the following draft recommended revisions to this section of the BRD.

- Additional discussion from the published literature about the advantages and limitations of using various supplemental metabolizing systems in cell culture for cytotoxicity testing could be included.
- Based on the Perloux et al. (1992) study, a discussion about whether the relatively good predictive value is a result of the route of exposure (intravenous [iv] and intraperitoneal [ip]), as well as information on the range of chemical types and the range of toxicities should be included. The poorer correlations for the oral route, along with the better correlations for the iv route, should be included. The correlation of different routes of exposure and the reflected kinetic variation should be discussed.
- The results of the workshop presented in Seibert et al. 1996 (Acute Toxicity Testing *In vitro* and the Classification and Labelling of Chemicals. The Report and Recommendations of ECVAM Workshop 16. Alternatives to Laboratory Animals 24:499-510) should be included.
- It would be useful to compare the range of *in vivo* toxicities and modes of action represented in these other studies reported in Section 9 with the present ICCVAM study.

- Clarification about the percentage reduction of animal use as referenced in the ICCVAM 2001a report should be included (i.e., what is the likely basis for the difference between then and now).

Dr. Stitzel asked for comments from the Panel on this section of the BRD. Since no comments were provided, the Panel agreed upon the draft recommended revisions.

BRD Section 10
Animal Welfare Considerations (Refinement, Reduction, and Replacement)

Dr. Rowan led the discussion of Section 10.
All supplemental data and information provided to the Panel via the NICEATM restricted website will be added to the final BRD.

Dr. Strickland stated that when the evaluation was performed with all of the reference substances, the RC millimole regression provided the best animal savings results, especially for substances with high toxicity. The Panel reviewed Table 1 from the AnimalUse.doc file provided on the restricted website. The biostatisticians questioned the difference in animal use for the default starting dose between the RC millimole regression and the other two regressions.

The Panel discussed whether or not a millimolar or a weight regression should be used to estimate the starting dose for acute oral toxicity tests. They recommended that if the molecular weight (MW) is unknown, the mg/kg regression should be used. If MW is known, they recommended using the mM regression since this would be more appropriate biologically. A decision tree may be needed to determine which regressions should be used for a test chemical. Other recommended revisions to this section of the BRD included:

- A substantial percent of the time the toxicity of "highly toxic" molecules *in vivo* was predicted to be less toxic using the cytotoxicity assays. In these instances, animals would be lost and subjected to untoward toxicities by using the higher predicted starting doses. Thus, the Panel recommended that the cytotoxicity tests only be used in a weight-of-evidence approach to determining starting doses for acute oral toxicity test methods.
- Although the accuracy appears to be low, it is still better than starting at the default starting dose if no other information is available.
- Based on existing data, where molecular weight information is available for a relatively pure test substance, the mM regression should be used; in the absence of such data, the mg/kg regression should be used.
- The possibility of using the NRU assays to determine the starting dose for the FDP acute toxicity test should be more carefully evaluated.
- Animal savings should take into account, to the extent possible, prevalence (i.e., the chemical distribution within the various GHS classifications).

Dr. Stitzel asked for comments from the Panel on this section of the BRD. Since no comments were provided, the Panel agreed upon the draft recommended revisions.

BRD Section 11
Practical Considerations

Dr. Ehrich provided a brief summary of Section 11 and listed the recommended revisions for the BRD.

The Panel agreed that extra efforts such as better education for laboratory technicians are needed for transferability of the test methods. Laboratories have their own ways of doing things and it is understandable to have differences in data. The protocols should have better detail to make sure everyone does the same thing during a test. The ICCVAM recommended performance standards and protocols should emphasize what education and proficiency is needed.

The Panel concluded that it is difficult to compare the value of the *in vitro* NRU assay per chemical to achieve an IC_{50} versus an animal test to achieve a LD_{50}. However, given that, the information presented in this section of BRD appears sufficient, with the following exceptions.

- It appears that transferability was not as easy as was stated. Minor protocol differences can have profound effects.
- Adequate training must be conducted prior to the initiation of the study.
- The costs for equipment and working time needed to perform the assays and a cost-benefit analysis should be included.
- NRU assays are not for replacement but for reduction. It would be appropriate to describe the reduction in the number of animals used.
- The time needed to prescreen NHK culture medium should be described.

Dr. Stitzel asked for comments from the Panel on this section of the BRD. Since no comments were provided, the Panel agreed upon the draft recommended revisions.

PUBLIC COMMENTS (Session 1)

Dr. Manfred Liebsch - Centre for Documentation and Evaluation of Alternatives to Animal Experiments (ZEBET) - Germany

Dr. Liebsch stated that he represented the ECVAM Scientific Advisory Committee (ESAC) Shadow Panel on the ICCVAM Peer Review of *In Vitro* Acute Toxicity Test Methods. The Shadow Panel's purpose is to facilitate a transparent communication process between ICCVAM and ECVAM. He provided the following comments:

- Why were the following recommendations from the ICCVAM *In vitro* Workshop of 2000 not adequately considered: (1) immediate implementation of the ZEBET Registry of Cytotoxicity approach to estimate acute toxicity starting doses, and (2) development of a 2-3 year validation study using *in vitro* methods to replace rodent acute oral toxicity testing
- The study's objectives were partly conflicting in regard to validation of the RC prediction model
- The selection of test chemicals was inappropriate to achieve the main study goal

- The *in vitro* data on intralaboratory and interlaboratory variations should be related to other multi-centre studies using NRU assays
- Take into account the influence of variability of both *in vitro* and *in vivo* data (in particular in the very toxic range) on the accuracy of predictions obtained
- Explain the poor fits of the data to the combined laboratory 3T3 and NHK regressions
- Appropriately discuss the study outcome in relation to other studies
- Take into account the prevalence chemicals, with respect to toxicity, for calculations of animal savings (not predictive power)

Ms Jessica Sandler – People for the Ethical Treatment of Animals (PETA)

Ms Sandler spoke of her involvement in the 1990s with the EPA and The Johns Hopkins Center for Alternatives to Animal Testing to impress upon the organizations that Dr. Bjorn Ekwall's methodology using cell death was an alternative to animal testing. She expressed dismay in the lack of interest by both groups in following this avenue. She also stated that toxicity tests should apply to the species of concern and that animal tests do not protect humans. She was critical of ICCVAM for not following the ICCVAM *In Vitro* Workshop 2000 recommendations on accepting non-animal testing as replacements. She stated that she believes ICCVAM's congressional mandate requires it to focus on the replacement of animals in lethal dose testing. Ms Sandler's public statement is available on the ICCVAM/NICEATM website in pdf format at the URL link provided (http://iccvam.niehs.nih.gov/methods/invidocs/brdcomm.htm).

Dr. Rodger Curren – Institute for *In vitro* Sciences (IIVS)

Dr. Curren thanked the Panel for their reviews and enthusiasm. He provided the following comments:

- A more accurate assessment of the "accuracy" of the method would be to model the results using a chemical set which more closely matched the original Halle chemical distribution in the RC regression. The current set of chemicals is biased toward outliers.
- The calculations for "animals saved" would be more informative if the data used for modeling was more representative of the original Halle chemical distribution in the RC regression
- It would be more logical to use the closest default dose to the estimated LD_{50} as the starting dose than to follow the OECD protocols which say to use the next lower dose (of a set of predetermined doses) to the value estimated by the cytotoxicity assay
- Minor comments included: the human response to digoxin is much higher than the animal response; information on most components of the keratinocyte growth medium should be available to researchers; the difference in SLS sensitivity between the two cell types may be influenced by the presence or absence of serum in the medium; the variability between labs should be examined more carefully to determine whether it is biologically significant

Final Review of the BRD

Dr. Stitzel asked the Panel to review the recommended revisions for each BRD section, taking into account the public comments, and to decide if additional changes are necessary. If no changes were recommended, then the recommendations for that section of the BRD were considered to as final.

No changes were made to the draft recommendations for Sections 1, 2, 4, 5, 7, 8, 9, and 11.

Section 3: The Panel asked for additional discussion of and reaction to the public comments from Dr. Manfred Liebsch. Dr. Stokes stated that the validation study tried to maximize the use of chemicals that had human and rat toxicity data. ECVAM is reviewing the human lethal serum/blood concentrations (LC) data for future use. Despite Dr. Liebsch's assertions, validation of the RC regression was not an objective of the NICEATM/ECVAM validation study. Dr. Stokes said that these clarifications would be in the final report. No other comments were made and the draft recommendations for this section were accepted by the Panel.

Section 6: The Panel asked for additional discussion of and reaction to the public comments from Dr. Rodger Curren. Dr. Seibert indicated the test methods should be so reliable that they could be done around the world, but there is no established and accepted criterion for reliability. Dr. Elmore suggested a graphical analysis in which the data from each individual laboratory is compared with the laboratory mean to determine whether one laboratory is different from the others. Dr. Stokes said this analysis could be added to the final report. No other comments were made and the draft recommendations for this section were accepted as final by the Panel.

Section 10: The Panel recommended the addition of prevalence data based on the reference from Dr. Liebsch. The accuracy number needs to be corrected in the BRD so that it reflects the right regression (i.e., the RC). No other comments were made and the draft recommendations for this section were considered to be final.

Validation Status

Dr. Stitzel asked the Panel whether the test methods are valid and supported by the data. The Panel agreed that the test methods are valid as a weight-of-evidence approach for estimating starting dose. Although the test methods are useful, they are not necessary and should not be made obligatory. Additional clarity is needed on how to use the weight-of-evidence approach, but this may require additional data.

The Panel agreed to the following statement on the validation aspect of the test methods.
The Panel agrees that the applicable validation criteria have been adequately addressed for using these in vitro test methods in a weight-of-evidence approach to determine the starting dose for acute oral in vivo toxicity protocols.

DRAFT ICCVAM RECOMMENDATIONS FOR *IN VITRO* ACUTE TOXICITY TEST METHODS

Presentation of Draft ICCVAM Recommendations

Dr. Marilyn Wind presented the draft ICCVAM recommendations for test method use and future studies. ICCVAM draft recommendations are now presented at peer review meetings due to OMB requirements for peer review of the scientific information used as the basis for the recommendations. Dr. Stitzel reminded the Panel that the discussion was to determine whether the scientific data and information in the BRD supports the ICCVAM recommendations.

Are the Draft ICCVAM Recommendations on Proposed Usefulness/Limitations Supported by the BRD?

Dr. Marsman led the discussion. The Panel agreed to the following statements in response to the ICCVAM recommendations.

(1) "*The 3T3 and NHK NRU test methods are not sufficiently accurate to predict the acute oral toxicity of substances for the purposes of hazard classification (see Section 6 of BRD).*"
- The Panel agrees with this statement in that neither of the two basal cytotoxicity tests can be used as alternatives for the *in vivo* acute oral toxicity test for the purposes of hazard classification.
- In the BRD, the rat data was not all generated in accordance with Good Laboratory Practice (GLP) standards

(2) "*For the purposes of acute oral toxicity testing, the 3T3 and NHK NRU test methods may be used in a weight-of-evidence approach to determine the starting dose for the current acute oral in vivo toxicity protocols (i.e., the UDP and ATC).*"
- The Panel agrees that the *in vitro* test methods may be useful in a weight-of-evidence approach to determine the starting dose for acute oral *in vivo* toxicity protocols.
- Given its limited predictive capacity, however, it is unclear whether it will provide substantial weight in that decision.
- The overall accuracy is modest, and enhancement of accuracy through material selection (modular approach), model refinement, or tiered testing strategy should be pursued.

(3) *"Consistent with the U.S. Government Principles on the Use of Animals in Research, Testing, and Education (National Research Council 1996), and the U.S. Public Health Service Policy on Humane Care and Use of Laboratory Animals (PHS 2002)*[1]*, in vitro basal cytotoxicity test methods as part of a weight-of-evidence approach to estimate the starting dose for acute oral in vivo toxicity test methods should be considered and used where appropriate before testing is conducted using animals. For some types of substances, this approach will reduce the number of animals needed. In some testing situations, the approach may also reduce the numbers of animals that die or need to be humanely killed."*
- The Panel agrees.

(4) *"Substances with specific toxic mechanisms that are not expected to be active in 3T3 or NHK cells (e.g., those that are neurotoxic, cardiotoxic, interfere with energy utilization, or alkylate proteins and other macromolecules) will likely be underpredicted by these in vitro basal cytotoxicity test methods. Therefore, until such time as a more predictive testing approach is developed, the results from basal cytotoxicity testing with such substances may not be appropriate."*

- The Panel disagrees with elements of this statement; specific toxic mechanisms that the BRD stated are not expected to be active in 3T3 and NHK cells, such as "interference with energy utilization and alkylation of proteins and other macromolecules", are mechanisms of cytotoxic action and should be detectable with 3T3 and NHK cells.

(5) *"The regression formula used to determine starting doses should be the RC regression line [with IC_{50} values in µg/mL and LD_{50} values in mg/kg] developed with the RC chemicals using rat LD_{50} data only and excluding chemicals with mechanisms of action that are not expected to be active in in vitro basal cytotoxicity test methods."*

- The Panel does not agree with this statement.
- There was consensus among the Panel that the data contained within the BRD or the open literature were not sufficient to justify the exclusion of materials based on mechanism.
- It is not justified to (retrospectively) exclude substances because of assumed modes of toxic action *in vivo* and/or possible involvement of biotransformation reactions.
- The Panel recommends that ICCVAM consider convening a work group to explore mechanisms of action of acute toxicity, and how acquiring additional information on acute toxic mechanisms might be put into practice under acute toxicity testing.
- Although a modular approach to use of the model looks like it may be more reliable, the database is likely too small for most mechanisms of action to draw sound conclusions regarding strengths and limitations of the test methods with respect to chemical classes, mechanisms of toxicity, or physico-chemical properties. Given that it is likely that mode of action for a random source material would be unknown, it is unlikely that a modular approach based upon mechanism is a viable option. A better approach to validation would be one based on chemical class, implying similar mode of action.
- *The Panel recommends moving the last two comments to the ICCVAM recommended future studies section.*

(6) *"The performance of other in vitro basal cytotoxicity test methods that are based on similar scientific principles and that measure or predict the same biological response (i.e., basal cytotoxicity and the rat acute oral LD_{50} value,*

respectively) should be demonstrated to meet or exceed the accuracy and reliability of the 3T3 and NHK NRU test methods."

- The Panel agrees with this statement although the reliability of the methods in this study was not quite satisfying (e.g., interlaboratory reproducibility), the reproducibility of these methods (e.g., intralaboratory reproducibility) are modest, and the accuracy of these methods are poor.

(7) *"Compared to the NHK NRU test method, the 3T3 NRU test method appears to be less labor intensive and less expensive to conduct; therefore, the 3T3 NRU cytotoxicity test method is recommended for general use."*

- Some Panel members agreed in a general sense, but cautioned that one model may be preferred over the other based upon specific knowledge regarding known mechanisms of action (e.g., the rationale for the disparate results observed with aminopterin and digoxin). Other Panel members agreed with this statement because the use of continuous cell lines is more efficient, especially since the overall animal savings were relatively low.
- One Panel member noted that NHK NRU IC_{50} data have shown a better correlation with human LC_{50} values (R^2=0.62) than do rodent 3T3 NRU IC_{50} data (R^2=0.51) and better than rodent LD_{50} data correlates with human LC_{50} values (R^2=0.56) as reported by S. Casati et al. at the 5th World Congress in Berlin, 2005. It is important to remember that hazard assessment relates to the safety of humans, not rats.
- Based on costs of commercial keratinocytes, the NHK NRU assay may be cost-prohibitive.
- The proprietary nature of the composition of the NHK culture medium makes it impossible to assess the role differences in media composition may have had on the results.

Draft Recommended Test Method Limitations

The Panel recommended adding the following verbiage to the draft report.

- Colored substances (besides red substances) may absorb light in the optical density range of the NRU assay and would affect the test system.
- The BRD indicates that optimization to allow for testing of mixtures was being undertaken, yet no mixtures were used in fitting the regression curve. Given the limitations of the assays in accurately predicting materials of known or uncertain mechanisms, the testing of mixtures seems highly controversial.

Dr. Stitzel asked for comments from the Panel on these draft ICCVAM recommendations as to the proposed usefulness and limitations of the two *in vitro* cytotoxicity test methods. No additional comments were provided and the Panel agreed unanimously with the draft revisions to the ICCVAM recommendations.

Are the Draft ICCVAM Recommended Standardized Protocols Supported by the BRD?

Dr. Ehrich led the discussion on the protocols. The Panel agreed that the protocols are generally quite detailed and laboratory technicians should be able follow the procedures. The Panel recommended the following clarifications be added to the 3T3 and NHK NRU test method protocols:

Protocol Recommendations

- The rationale for testing the positive control on separate plates rather than on the test plates should be provided.
- The number of definitive tests that should be performed for a test substance should be specified.
- The range of linearity of the microplate reader should be confirmed (as per in-house Standard Operation Procedures [SOPs]) for the recommended optical density (OD_{540}) and stated.
- Maximum absorbance values needed by a spectrophotometric plate reader should be provided for application to the NRU assays.
- The test method protocols should be streamlined. (Undefined as to how this should be accomplished.)
- Guidance for using methods other than the Hill function to determine IC_{50} values should be provided.
- The lowest acceptable test substance dilution factor (i.e., 1.21) should be reduced rather than accepting only one cytotoxicity point between 0 and 100% viability on a steep dose-response curve to use for determination of the IC_{50} value.
- Study directors and quality assurance units are necessary only if testing is performed under GLP.
- Good cell culture practices (e.g., Hartung et al. 2002) must be followed.
- Whether or not a prequalification test of new keratinocytes should be performed by the laboratory prior to actual testing should be stated.
- A recommendation that keratinocytes should be procured only through commercial sources and not by preparing primary cultures from donated tissue should be included.
- Additional guidance to the solubility step-wise procedure should be added (i.e., ensure that test substance solution preparation procedures can be easily understood by laboratory technicians).
- The need for training of laboratory technicians so they may be able to better understand solvent and solubility determinations should be included.
- Additional guidance as to the use of a microscope to assist in determining solubility of a test substance should be added.
- Test substances that may etch plastic or “film out” in medium should be identified (the importance of detecting such compounds by the laboratory technicians should be emphasized).
- The protocols should recommend the use of a solvent (e.g., dimethyl sulfoxide [DMSO], ethanol) at its lowest possible concentration at each test substance concentration level.

- There is concern about the differences in solvent selection between laboratories as compared to the BioReliance solvent information. The variability between laboratories in the selection of solvent points out a possible flaw in the solvent determination protocol. This should be evaluated for future studies.

Dr. Stitzel asked for comments from the Panel on these draft ICCVAM standardized protocols for the two *in vitro* cytotoxicity test methods. Since no additional comments were provided. The Panel agreed unanimously with the draft recommended revisions to the draft ICCVAM standardized protocols.

Are the Draft ICCVAM Recommended Test Method Performance Standards Supported by the BRD?

Dr. Elmore presented the Panel comments on whether the ICCVAM draft recommended test method performance standards were supported by the BRD.

The available data from this study appear to support the validity of the recommended performance standards for the test methods. The usefulness and limitations are well covered, and if validated, the methods may be a worthwhile option. However, there may be some cause for concern if use of the methods is made compulsory for regulatory purposes.

- Recommendations made in section 2.3.2 (Application of the Test Substances), section 2.3.3 (Control Substances), and section 2.3.4 (Viability Measurements) are acceptable.
- A discussion is needed about whether or not the NRU assays are recommended for use with unknown substances and mixtures.
- The significance of the secondary chemical subset to be used for "investigational purposes" should be better elucidated in the document.

Dr. Stitzel asked for discussion from the Panel on whether the draft ICCVAM recommended performance standards for the two *in vitro* cytotoxicity test methods were supported by the BRD. No additional comments were provided. The Panel agreed unanimously with the draft recommended revisions to the ICCVAM recommendations.

Are the Draft ICCVAM Recommended Future Studies Supported by the BRD?

Dr. Rowan presented the Panel comments on whether the ICCVAM draft recommendations on the recommended future studies were supported by the BRD. He stated that efforts should be made to collect GLP LD_{50} data from industry for use in *in vitro/in vivo* databases. The ICCVAM recommendations were discussed and the bullets below represent the Panel's responses.

(1) ICVAM draft recommendation: *"Additional data should be collected using the 3T3 and/or the NHK NRU test methods to evaluate their usefulness for predicting the in vivo acute oral toxicity of chemical mixtures."*

- The Panel generally agrees that this is a good recommendation, although collecting data could be difficult and doing correlation with *in vivo* data would be even more difficult. It may be useful to suggest that such data

only be collected with the 3T3 NRU test method, and that it would be necessary to clarify the reasons for the interlaboratory variations for future use of the method.

(2) ICVAM draft recommendation: *"Additional high quality comparative in vitro basal cytotoxicity data should be collected in tandem with in vivo rat acute oral toxicity test results to supplement the high quality validation database started by this study. Periodic evaluations of the expanded database should be conducted to further characterize the usefulness and limitations of using in vitro cytotoxicity data as part of a weight-of-evidence approach to estimate starting doses."*

- The Panel believes this could be valuable under certain conditions, especially if NRU data are collected as acute toxicity testing is conducted.
- However, no panel member wants *in vivo* testing conducted solely to collect data to assess the usefulness of the NRU test, particularly given that the savings in animal numbers that arise from the use of the NRU test to determine the starting dose for the ATC method or UDP are fairly modest.

(3) ICVAM draft recommendation: *"Additional efforts should be conducted to identify additional in vitro tests and other methods necessary to achieve accurate acute oral hazard classification; specifically, studies should be conducted to investigate the potential use of in vitro cell-based test methods that incorporate mechanisms of action and evaluations of ADME to provide improved estimates of acute toxicity hazard categories."*

- The Panel agrees with this statement and adds that there should be additional effort towards development of alternative methods to adequately predict the *in vivo* acute toxicity of chemicals for the purposes of hazard classification.
- An additional statement to include could be, "and the development of methods to extrapolate from *in vitro* toxic concentrations to equivalent doses *in vivo*."

(4) ICVAM draft recommendation: *"The in vivo database of reference substances used in this validation study should be used to evaluate the utility of other non-animal approaches to estimate starting doses for acute oral systemic toxicity tests (e.g., widely available software that uses quantitative structure-activity relationships [QSAR])."*

- The Panel agreed with this recommendation.

(5) ICVAM draft recommendation: *"Standardized procedures to collect information pertinent to an understanding of the mechanisms of lethality should be included in future in vivo rat acute oral toxicity studies. Such information will likely be necessary to support the further development of predictive mechanism-based in vitro methods."*

- The Panel agrees with this recommendation; this is really important and could further the development of non-animal alternatives in the future.
- To facilitate comparisons and model development, future studies should incorporate high quality animal data for required testing of new agents, (where possible) blood levels from animals (LC_{50}), and high quality *in vitro* data from the same agents.
- The Panel recommends that ICCVAM consider convening a work group to identify the appropriate *in vivo* endpoints to assess during acute toxicity testing so as to generate information on mechanisms of acute toxicity.
- Although a modular approach to use of the model looks like it may be more reliable, the data base is likely too small for most mechanisms of action to draw sound conclusions regarding strengths and limitations of the test methods with respect to chemical classes, mechanisms of toxicity, or physico-chemical properties. Given that it is likely that a mode of action is unlikely to be known about a random source material, it is unlikely that a modular approach based upon mechanism is often going to be a viable option. A more likely approach to validation would be one based on chemical class, implying similar mode of action.

(6) ICVAM draft recommendation: *"An expanded list of reference substances with estimated rat LD_{50} values substantiated by high quality in vivo data should be developed for use in future in vitro test method development and validation studies."*

- The Panel agrees with this recommendation; there should be a concerted effort to collect proprietary data.

Dr. Stitzel asked for comments from the Panel on these draft ICCVAM recommendations for future studies. Since no additional comments were provided, the Panel agrees with the draft revisions to the ICCVAM recommendations.

PUBLIC COMMENTS (Session 2)

Ms Kristie Stoick - Physicians Committee for Responsible Medicine (PCRM)

Ms Stoick introduced herself as a representative of the PCRM and requested that a full replacement of *in vivo* testing be sought. She appreciates the progress toward reduction and refinement of animal use in acute toxicity evaluations, but suggests that total replacement, rather than reduction and refinement, is the solution to poor concordance. She faulted ICCVAM for not following up on the research and development recommendations from the ICCVAM *In Vitro* Workshop in 2000. She expressed hope that the appropriate government agencies will implement any validated reduction and refinement measures and urges the implementation of a dedicated mechanism to collect all data generated from these tests for evaluation and determination of its usefulness in replacing *in vivo* acute toxicity tests.

Final Review of the Draft ICCVAM Recommendations

Dr. Stitzel asked if any Panel member wanted make any changes to the comments of the Panel regarding the draft ICCVAM test method recommendations. No further changes were requested. Dr. Stitzel affirmed that the Panel unanimously concurred with all of the above comments. The Panel agreed also that the statement on validation of the test methods was acceptable.

Concluding Remarks

Drs. Stitzel and Stokes thanked the Panel members for their time and effort.

Adjournment

The meeting was adjourned at 5:23 p.m.

William S. Stokes, D.V.M.

NIEHS

P.O. Box 12233

MD-EC17

Research Triangle Park, NC 27709

Dear Dr. Stokes:

The Meeting Summary, Peer Review Panel Public Meeting, *In Vitro* Methods for Estimating Starting Doses for Acute Systemic Toxicity Testing, dated May 23, 2006, accurately summarizes the Peer Review Panel Public meeting of May 23, 2006, in Bethesda, MD.

Sincerely,

| --- | --- | --- |
| Signature | Printed Name | Date |

APPENDIX B
RELEVANT FEDERAL ACUTE ORAL TOXICITY REGULATIONS AND TESTING GUIDELINES

[This Page Intentionally Left Blank]

APPENDIX B1

TABLE OF RELEVANT ACUTE ORAL TOXICITY REGULATIONS

(Note to the Reader: Regulations may be updated in the future. It is recommended that users review the most current version of all regulations identified. Electronic versions of the regulations can be obtained at: http://www.gpoaccess.gov/nara/index.html)

[This Page Intentionally Left Blank]

AGENCY	TITLE	CHAPTER	PART AND TITLE	SECTION	
CPSC	16	II	PART 1500--HAZARDOUS SUBSTANCES AND ARTICLES; ADMINISTRATION AND ENFORCEMENT REGULATIONS	1500.3	Definitions
DOT	49	I	PART 173--SHIPPERS--GENERAL REQUIREMENTS FOR SHIPMENTS AND PACKAGINGS	173.132	Class 6, Division 6.1 - Definitions
				173.133	Assignment of Packing Group and Hazard Zones for Divusion 6.1 Materials
EPA	40	I	PART 156--LABELING REQUIREMENTS FOR PESTICIDES AND DEVICES	156.10	Labeling Requirements
				156.620	Toxicity Category.
EPA	40	1	157: PACKAGING REQUIREMENTS FOR PESTICIDES AND DEVICES	157.22	When required.
EPA	40	1	158: DATA REQUIREMENTS FOR REGISTRATION	158.202	Purposes of the registration data requirements.
				158.340	Toxicology data requirements.
				158.690	Biochemical pesticides data requirements.
				158.740	Microbial pesticides--Product analysis data requirements.
EPA	40	I	159: STATEMENTS OF POLICIES AND INTERPRETATIONS	159.165	Toxicological and ecological studies.
OSHA	29	XVII	1910: OCCUPATIONAL SAFETY AND HEALTH STANDARDS	1910.1200	Hazard communication.

Abbreviations: CPSC=U.S. Consumer Products Safety Commission; DOT=Department of Transportation; EPA=U.S. Environmental Protection Agency; OSHA=Occupational and Safety Administration.

[This Page Intentionally Left Blank]

APPENDIX B2

OECD GUIDELINE 425: ACUTE ORAL TOXICITY – UP-AND-DOWN PROCEDURE

[This Page Intentionally Left Blank]

APPENDIX B3

OECD GUIDELINE 423: ACUTE ORAL TOXICITY – ACUTE TOXIC CLASS METHOD

[This Page Intentionally Left Blank]

APPENDIX B4
OECD GUIDELINE 420: ACUTE ORAL TOXICITY – FIXED DOSE PROCEDURE

[This Page Intentionally Left Blank]

APPENDIX B5

Health Effects Test Guidelines OPPTS 870.1100: Acute Oral Toxicity

[This Page Intentionally Left Blank]

APPENDIX C

RECOMMENDED TEST METHOD PROTOCOLS

[This Page Intentionally Left Blank]

APPENDIX C1

TEST METHOD PROTOCOL FOR THE BALB/c 3T3 NRU CYTOTOXICITY TEST METHOD

[This Page Intentionally Left Blank]

ICCVAM Recommended Protocol for the BALB/c 3T3 Neutral Red Uptake (NRU) Cytotoxicity Test - A Test for Basal Cytotoxicity

1.0 PURPOSE

This test method is used to evaluate the cytotoxicity of test substances using the BALB/c 3T3 Neutral Red Uptake (NRU) *in vitro* cytotoxicity test. The data generated from the *in vitro* cytotoxicity assays are used to predict the starting doses for rodent acute oral systemic toxicity assays. This test method protocol outlines the procedures for performing the basal cytotoxicity test and is the result of the joint independent *in vitro* validation study organized by National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the European Centre for the Validation of Alternative Methods (ECVAM).

If changes or modifications are made to this protocol, the testing laboratory should prove that the results are comparable to those obtained when using this original protocol.

2.0 TEST SYSTEM

The NRU cytotoxicity assay procedure is based on the ability of viable cells to incorporate and bind neutral red (NR), a supravital dye. NR is a weak cationic dye that readily diffuses through the plasma membrane and concentrates in lysosomes where it electrostatically binds to the anionic lysosomal matrix. Toxicants can alter the cell surface or the lysosomal membrane to cause lysosomal fragility and other adverse changes that gradually become irreversible. Thus, cell death and/or inhibition of cell growth decreases the amount of neutral red retained by the culture. Healthy proliferating mammalian cells, when properly maintained in culture, continuously divide and multiply over time. A toxic chemical, regardless of site or mechanism of action, will interfere with this process and result in a reduction of the growth rate as reflected by cell number.

Cytotoxicity is expressed as a concentration dependent reduction of the uptake of NR after chemical exposure, thus providing a sensitive, integrated signal of both cell integrity and growth inhibition.

3.0 KEY PERSONNEL

3.1 Laboratory

- Study Director (only recommended if testing is performed under Good Laboratory Procedures [GLP])
- Laboratory Technician(s)

3.2 Testing Facility

- Scientific Advisor
- Quality Assurance Director (only necessary if testing is performed under GLP)
- Safety Manager
- Facility Management

4.0 DEFINITIONS

Hill function: a four parameter logistic mathematical model relating the concentration of test substance to the response being measured in a sigmoidal shape.

$$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(\log EC_{50} - X)\text{HillSlope}}}$$

where Y = response, X = the logarithm of dose (or concentration), Bottom = the minimum response, Top the maximum response, $\log EC_{50}$ = logarithm of X at the response midway between Top and Bottom, and HillSlope = the steepness

of the curve. When Top = 100 and Bottom = 0, the EC_{50} is the concentration at 50% viability (i.e., the IC_{50})

Documentation: all methods and procedures will be noted in a study workbook; logs will be maintained for general laboratory procedures and equipment (e.g., media preparation, test substance preparation, incubator function); all optical density data obtained from the spectrophotometer plate reader will be saved in electronic and paper formats; all calculations of ICx values and other derived data will be in electronic and paper format; all data will be archived

IC_{50}: test substance concentration producing 50% inhibition of the endpoint measured (i.e., cell viability)

5.0 IDENTIFICATION OF CONTROL SUBSTANCES

5.1 Positive Control (PC)

- Sodium Lauryl Sulfate (SLS)

5.2 Vehicle Control (VC)

- Assay medium (Dulbecco's Modification of Eagle's Medium [DMEM] containing 5% New Born Calf Serum (NCS), 4 mM L-Glutamine, 100 IU/mL Penicillin, 100 µg/mL Streptomycin)

5.3 Solvent Control

- VC control with solvent (i.e., assay medium, dimethyl sulfoxide [DMSO], or ethanol [ETOH]). DMSO is the preferred solvent for substances that are not water (i.e., assay medium) soluble.

6.0 PROCEDURES

6.1 Materials

6.1.1 Cell Line

- BALB/c 3T3 cells, clone A31 (e.g., CCL-163, American Type Culture Collection [ATCC], Manassas, VA, USA)

6.1.2 Technical Equipment[4]

- Incubator: 37 ºC ± 1 ºC, 90% ± 10% humidity, 5.0% ± 1.0% CO_2/air
- Laminar flow clean bench/cabinet (standard: "biological hazard")
- Waterbath: 37 ºC ± 1 ºC
- Inverse phase contrast microscope
- Sterile glass tubes with caps (e.g., 5 mL)
- Centrifuge
- Laboratory balance
- 96-well plate spectrophotometer (i.e., plate reader) equipped with 540 nm ± 10 nm filter with maximum absorbance of 3
- Shaker for microtiter plates
- Cell counter or hemocytometer
- Pipetting aid
- Pipettes, pipettors (multi-channel and single channel; multichannel repeater pipette), dilution block
- Cryotubes
- Tissue culture flasks (e.g., 75 - 80 cm^2, 25 cm^2)
- 96-well flat bottom tissue culture microtiter plates (e.g., Nunc # 167 008; Falcon tissue culture-treated)
- pH paper (wide and narrow range)
- Multichannel reagent reservoir

[4] Suggested brand names/vendors are listed in parentheses. Equivalents may be used.

- Waterbath sonicator
- Magnetic stirrer
- Antistatic bar ionizer/antistatic gun (optional for neutralizing static on 96-well plates)
- Dry heat block (optional)
- Adhesive film plate sealers (e.g., Excel Scientific SealPlate™, Cat # STR-SEAL-PLT or equivalent)
- Vortex mixer
- Filters/filtration devices

Note: Prescreen tissue culture flasks and microtiter plates to ensure that they adequately support the growth of 3T3 cells. Use multi-channel repeater pipettes for plating cells in the 96-well plates, dispensing plate rinse solutions, NR medium, and desorb solution. Do not use the repeater pipette for dispensing test substances to the cells.

6.1.3 Chemicals, Media, and Sera

- DMEM without L-Glutamine; should have high glucose [4.5 g/L] (e.g., ICN-Flow Cat. No. 12-332-54)
- L-Glutamine 200 mM (e.g., ICN-Flow # 16-801-49)
- NCS (e.g., Biochrom # SO 125)
- 0.05% Trypsin/0.02% Ethylenediaminetetraacetic acid (EDTA) solution (e.g., SIGMA T 3924, ICN-Flow, # 16891-49)
- Phosphate buffered saline (PBS) without Ca^{2+} and Mg^{2+}(for trypsinization)
- Hanks' Balanced Salt Solution (HBSS) without Ca^{2+} and Mg^{2+}(CMF-HBSS)
- Dulbecco's Phosphate Buffered Saline (D-PBS) for rinsing (formulation containing calcium and magnesium cations; glucose optional)
- Penicillin/streptomycin solution (e.g. ICN-Flow # 16-700-49)
- NR Dye – tissue culture-grade; liquid form (e.g., SIGMA N 2889); powder form (e.g., SIGMA N 4638)

- DMSO, U.S.P. analytical grade (Store under nitrogen @ -20 ºC)
- ETOH, U.S.P. analytical grade (100%, non-denatured for test substance preparation; 95% can be used for the desorb solution)
- Glacial acetic acid, analytical grade
- Distilled H_2O or any purified water suitable for cell culture and NR desorb solution (sterile)
- Sterile/non-sterile paper towels (for blotting 96-well plates)

Note: Due to lot variability of NCS, first check a lot for growth stimulating properties with 3T3 cells (approximately 20-24 hours doubling time) and then reserve a sufficient amount of NCS.

6.2 Preparation of Media and Solutions

Note: All solutions (except NR stock solution, NR medium and NR desorb), glassware, pipettes, etc., shall be sterile and all procedures should be carried out under aseptic conditions and in the sterile environment of a laminar flow cabinet (biological hazard standard). All methods and procedures will be adequately documented.

6.2.1 Media

DMEM (buffered with sodium bicarbonate) supplemented with (final concentrations in DMEM are quoted):

- **Freeze Medium:** contains 2X concentration of NCS and DMSO of final freezing solution
  - 40% NCS
  - 20% DMSO
- **Routine Culture Medium**
  - 10% NCS
  - 4 mM L-Glutamine

- **Chemical Dilution Medium[5]**
  - 4 mM L-Glutamine
  - 200 IU/mL Penicillin
  - 200 µg/mL Streptomycin
- **NR Dilution Medium**
  - 5% NCS
  - 4 mM Glutamine
  - 100 IU/mL Penicillin
  - 100 µg/mL Streptomycin

Completed media formulations should be kept at approximately 2-8 °C and stored for no longer than two weeks.

6.2.2 NR Stock Solution

- The liquid tissue culture-grade stock NR Solution is the first choice (e.g., SIGMA #N2889, 3.3 mg/mL). Store liquid tissue culture-grade NR Stock Solution at the storage conditions and shelf-life period recommended by the manufacturer.
- A stock solution can be made with powder NR dye and water (e.g., 0.25 g NR Dye powder in 100 mL H_2O) if the liquid stock form is not available. The stock should be stored in the dark at room temperature for up to two months.

6.2.3 NR Medium

EXAMPLE:

0.758 mL (3.3 mg NR dye/mL sol.)	NR Stock Solution
99.242 mL	NR Dilution Medium (pre-warmed to 37 °C)

[5] The Chemical Dilution Medium with test substance will dilute the serum concentration of the Routine Culture Medium in the test plate to 5%. Serum proteins may mask the toxicity of the test substance, but serum cannot be totally excluded because cell growth is markedly reduced in its absence.

The final concentration of the NR Medium is 25 µg NR dye/mL and aliquots will be prepared on the day of application.

Note: Filter the NR Medium (e.g., Millipore filtering, 0.2 – 0.45 µm pore size) to reduce NR crystals. Maintain aliquots of the NR Medium at 37 °C (e.g., in a waterbath) before adding to the cells and use within 60 minutes of preparation and within 15 minutes after removing from 37 °C storage. Examine the solution for crystals prior to use.

6.2.4 ETOH/Acetic Acid Solution (NR Desorb)

- 1% Glacial acetic acid solution
- 50% ETOH
- 49% H_2O

6.3 Methods

6.3.1 Cell Maintenance and Culture Procedures

- BALB/c 3T3 cells are routinely grown as a monolayer in tissue culture grade flasks (e.g., 75 - 80 cm^2) at 37 ºC ± 1 ºC, 90% ± 10% humidity, and 5.0% ± 1.0% CO_2/air.
- Examine the cells on a daily (i.e., on workdays) basis under a phase contrast microscope, and note any changes in morphology or their adhesive properties in a study workbook.
- All cell culture studies should follow good cell culture practices (Hartung et al. 2002).

6.3.2 Receipt of Cryopreserved BALB/c 3T3 Cells

Upon receipt of cryopreserved BALB/c 3T3 cells, store the vial(s) of cells in a liquid nitrogen freezer until needed.

6.3.3 Thawing Cells

Thaw a fresh batch of frozen cells from the stock lot of cells and culture approximately every two months. This period resembles a sequence of about 18 passages.

- Thaw cells by putting ampules into a waterbath at 37 °C ± 1 ºC. Leave for as brief a time as possible.
- Resuspend the cells in pre-warmed Routine Culture Medium and transfer into pre-warmed Routine Culture Medium in a tissue-culture flask.
- Incubate at 37 ºC ± 1 ºC, 90% ± 10% humidity, and 5.0% ± 1.0% CO_2/air.
- When the cells have attached to the bottom of the flask (within 4 to 24 hours), decant the supernatant and replace with fresh pre-warmed (37 ºC) medium. Culture as described above.
- Passage at least two times before using the cells in a cytotoxicity test.

6.3.4 Routine Culture of BALB/C 3T3 Cells

Remove cells from the flask by trypsinization when they exceed 50% confluence (but less than 80% confluent):

- Decant medium, briefly rinse cultures with 5 mL PBS or HBSS (without Ca^{2+}, Mg^{2+}) per 25 cm^2 flask (15 mL per 75 cm^2 flask). Wash cells by gentle agitation to remove any remaining serum that might inhibit the action of the trypsin.
- Discard the washing solution. Repeat the rinsing procedure and discard the washing solution.
- Add 1-2 mL trypsin-EDTA solution per 25 cm^2 to the monolayer for a few seconds (e.g., 15-30 seconds).
- Remove excess trypsin-EDTA solution and incubate the cells at room temperature.
- After 2-3 minutes, lightly tap the flask to detach the cells into a single cell suspension.

6.3.5 Cell Counting

- After the cells are detached, add 0.1-0.2 mL of pre-warmed (37 ºC) Routine Culture Medium/cm^2 to the flask (e.g., 2.5 mL for a 25 cm^2 flask).
- Disperse the monolayer by gentle trituration to obtain a single cell suspension for exact counting.
- Count a sample of the cell suspension obtained using a hemocytometer or cell counter (e.g., Coulter counter).

6.3.6 Subculture of Cells

BALB/c 3T3 cells are routinely sub-cultured into other flasks or seeded into 96-well microtiter plates (see **Figure C1-1** for 96-well test plate configuration) and passaged at suggested cell densities as listed in **Table C1-1** (approximate doubling time is 20-24 hours). Laboratories must determine and adjust the final density to achieve appropriate growth.

Figure C1-1 96-Well Plate Configuration for Positive Control (PC) and Test Substance Assays

	1	2	3	4	5	6	7	8	9	10	11	12
A	VCb	VCb	C_1b	C_2b	C_3b	C_4b	C_5b	C_6b	C_7b	C_8b	VCb	VCb
B	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
C	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
D	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
E	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
F	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
G	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
H	VCb	VCb	C_1b	C_2b	C_3b	C_4b	C_5b	C_6b	C_7b	C_8b	VCb	VCb

VC1 and VC2 = VEHICLE CONTROL
C_1 – C_8 = Test Substances or PC (SLS) at eight concentrations (C_1 = highest, C_8 = lowest)
C_xb = BLANKS (Test substance or PC, but contain **no** cells)
VCb = VEHICLE CONTROL BLANK (contain **no** cells)

Table C1-1 Cell Density Guidelines for Subculturing

Days in Culture	Seeding Density (cells/cm^2)	Total Cells per 25 cm^2 flask	Total Cells per 75 cm^2 flask
2	16800	4.2×10^5	1.26×10^6
3	8400	2.1×10^5	6.3×10^5
4	4200	1.05×10^5	3.15×10^5

Note: It is important that cells have overcome the lag growth phase when they are used.

6.3.7 Freezing Cells (procedure required only if current stock of cells is depleted)

Store stocks of BALB/c 3T3 cells in sterile freezing tubes in a liquid nitrogen freezer. DMSO is used as a cryoprotective agent.

- Centrifuge trypsinized cells at approximately 200 x *g*.
- Suspend the cells in cold Routine Culture Medium (half the final freezing volume) so a final concentration of 1-5x10^6 cells/mL can be attained.
- Slowly add cold Freeze Medium to the cells so that the solvent will equilibrate across the cell membranes. Bring the cell suspension to the final freezing volume. The final cell suspension will be 10% DMSO. Aliquot the cell suspension into freezing tubes and fill to 1.8 mL.
- Place the tubes into an insulated container (e.g., styrofoam trays) and place in a freezer (-70 to -80 °C) for 24 hours (~freezing rate of 1 °C/minutes). The laboratory needs to ensure that the freezing protocol is applicable to the 3T3 cells and that the cells are viable when removed from cryopreservation.
- Place the frozen tubes into liquid nitrogen for storage.

6.3.8 Preparation of Cells for Assays

- Cultured cells that will be used in seeding the 96-well plates should be fed fresh medium the day before subculturing to the plates.[6]

[6]Note the seeding density to ensure that the cells in the control wells are not overgrown after three days (i.e., 24 hour incubation and 48 hour exposure to test substances). Prepare one plate per substance to be tested.

- Prepare a cell suspension of 2.0 – 3.0x10^4 cells/mL in Routine Culture Medium on the day of plate seeding.
- Use a multi-channel pipette to dispense 100 µL Routine Culture Medium only into the peripheral wells (blanks) of a 96-well tissue culture microtiter plate (See **Figure C1-1**).
- Dispense 100 µL of a cell suspension of 2.0 – 3.0x10^4 cells/mL (= 2.0 – 3.0x10^3 cells/well) in the remaining wells.

- Incubate cells for 24 hours ± 2 hours (37 ºC ± 1 ºC, 90% ± 10% humidity, 5.0% ± 1.0% CO_2/air) so that cells form a less than half (<50%) confluent monolayer. This incubation period assures cell recovery and adherence and progression to exponential growth phase.
- Examine each plate under a phase contrast microscope to assure that cell growth is relatively even across the microtiter plate. This check is performed to identify experimental and systemic cell seeding errors. Record observations in the study workbook.

6.3.9 Determination of Doubling Time

- Establish cells in culture and trypsinize cells as per **Section 6.3.4** for subculture. Resuspend cells in NR Dilution Medium (5% NCS). Seed cells at 4200 cells/cm^2.
- Seed five sets of cell culture vessels in triplicate (e.g., 15 tissue culture dishes [60mm x 15mm]). Use appropriate volume of culture medium for the culture vessels. Note number of cells placed into each culture dish. Place dishes into the incubators (37 ºC ± 1 ºC, 90% ± 10% humidity, 5.0% ± 1.0% CO_2/air).
- After 4 - 6 hours (use the same initial measurement time for each subsequent doubling time experiment), remove three culture dishes and trypsinize cells.
- Count cells using a cell counter or hemocytometer and document. Study Director may determine cell viability by dye exclusion (e.g., Trypan Blue; Nigrosin). Use appropriate size exclusion limits if using a Coulter counter.

- Repeat sampling at 24-, 48-, 72-, and 96-hours post inoculation. Change culture medium at 72 hours or sooner in remaining dishes if indicated by pH drop.
- Plot cell concentration (per mL of medium) on a log scale against time on a linear scale. Determine lag time and population doubling time. Additional dishes and time are needed if the entire growth curve is to be determined (lag phase, log phase, plateau phase).

6.4 Preparation of Test Substances

Note: Preparation under red or yellow light is recommended to preserve substances that degrade upon exposure to light.

Test substance solubility should be determined by following the procedures outlined in **ANNEX I** of this protocol.

6.4.1 Test Substances in Solution

- Equilibrate test substances to room temperature before dissolving and diluting.
- Prepare test substance immediately prior to use rather than preparing in bulk for use in subsequent tests. Ideally, the solutions must not be cloudy nor have noticeable precipitate. Each stock dilution should have at least 1-2 mL total volume to ensure adequate solution for the test wells in a single 96-well plate. The Study Director may store an aliquot (e.g., 1 mL) of the highest 2X stock solution (e.g., low solubility substances) in a freezer (e.g., -70 °C) for use in future substance analyses.
- For substances dissolved in DMSO or ETOH, the final DMSO or ETOH concentration for application to the cells must be 0.5% (v/v) in the vehicle controls and in all of the eight test concentrations. The concentration of DMSO or ETOH should be at the lowest possible concentration needed to dissolve the test substance.

- The stock solution for each test substance should be prepared at the highest concentration found to be soluble in the solubility test conducted per **ANNEX I**. Thus, the highest test concentration applied to the cells in each range finding experiment is:
  - 0.5 times the highest concentration found to be soluble in the solubility test, if the substance was soluble in Chemical Dilution Medium, or
  - 1/200 the highest concentration found to be soluble in the solubility test if the substance was soluble in ETOH or DMSO.

Example: Preparation of Test Substance in Solvent for Range Finding Experiments Using a Log Dilution Scheme

If DMSO is determined to be the preferred solvent at Tier 3 of the solubility test (i.e., 200,000 μg/mL), dissolve the substance in DMSO at 200,000 μg/mL for the chemical stock solution. The seven lower concentrations in the range finding experiment are prepared by successive dilutions that decrease by one log unit each.

- Label eight tubes 1 – 8. Add 0.9 mL solvent (e.g., DMSO) to tubes 2 - 8.
- Prepare stock solution of 200,000 μg test substance/mL solvent in tube # 1.
- Add 0.1 mL of 200,000 μg/mL dilution from tube #1 to tube #2 to make a 1:10 dilution in solvent (i.e., 20,000 μg/mL).
- Add 0.1 mL of 20,000 μg/mL dilution from tube #2 to tube #3 to make another 1:10 dilution (i.e., 1:100 dilution from stock solution) in solvent (i.e., 2,000 μg/mL). Continue making serial 1:10 dilutions in the prepared solvent tubes.
- Since each concentration is 200 fold greater than the concentration to be tested, make a 1:100 dilution by diluting 1 part dissolved test substance in each tube with 99 parts of Chemical Dilution Medium (e.g., 0.1 mL test substance in DMSO + 9.9 mL Chemical Dilution Medium) to derive the eight 2X concentrations for application to 3T3 cells. Each 2X test substance concentration will then contain 1% (v/v) solvent.

- The 3T3 cells will have 0.05 mL Routine Culture Medium in the wells prior to application of the test substance. By adding 0.05 mL of the appropriate 2X test substance concentration to the appropriate wells, the test substance will be diluted appropriately (e.g., highest concentration in well will be 1,000 µg/mL) in a total of 0.1 mL and the solvent concentration in the wells will be 0.5% (v/v).

- A test substance prepared in Chemical Dilution Medium, DMSO, or ETOH may precipitate upon transfer into the Routine Culture Medium. The 2X dosing solutions should be evaluated for precipitates and the results recorded in the study workbook. It is permissible to test all of the dosing solutions in the dose range finder experiments and main experiments. However, doses containing test substance precipitates should be avoided because it creates doubt about the concentration of test substance exposed to the cells.

Document all test substance preparations in the study workbook.

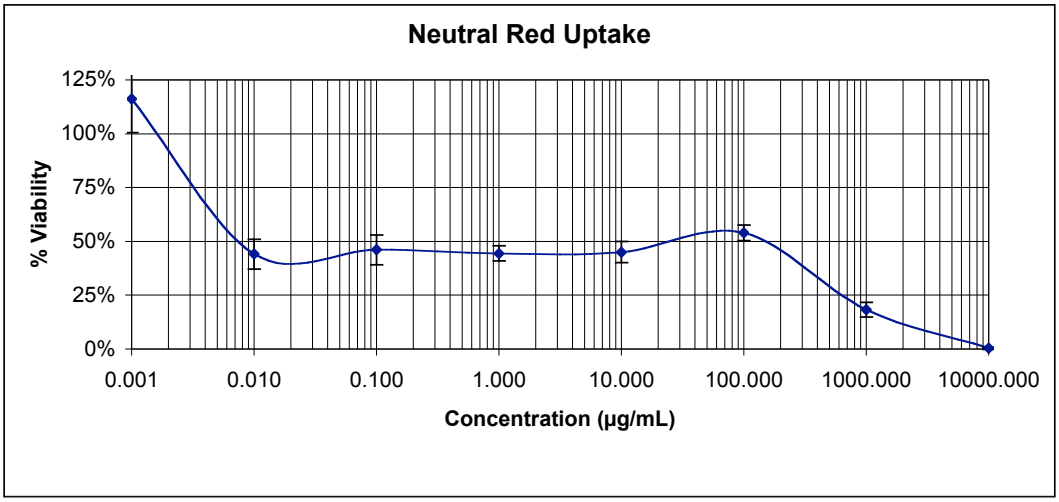
6.4.2 pH of Test Substance Solutions

- Prior to or immediately after application of the test substance to the 96-well plate, measure the pH of the highest 2X dosing concentration of the test substance (i.e., C1 in the test plate, see **Figure C1-1**) in culture medium.
- Use pH paper (e.g., pH 0 - 14 to estimate and pH 5 – 10 to determine more precise value; or Study Director's discretion) for measurements. The pH paper should be in contact with the solution for approximately one minute.
- Document the pH and note the color of the 2X concentration medium (i.e., in the Microsoft Excel® template; see **ANNEX II** for an example template). Medium color for all dosing dilutions should be noted in the study workbook. Do not adjust the pH.

6.4.3 Concentrations of Test Substance

- Range Finder Experiment
  - Test eight concentrations of the test substance by diluting the stock solution using log dilutions (e.g., 1:10, 1:100, 1:1000, etc.).
  - If a range finder experiment does not generate enough cytotoxicity, then higher doses should be attempted. If cytotoxicity is limited by solubility, then more stringent solubility procedures to increase the stock concentration (to the maximum concentration specified in **Section 6.4.4**) should be employed.
  - Place the test substance concentration into an incubator (37 ºC ± 1 ºC, 90% ± 10% humidity, 5.0% ± 1.0% CO_2/air) and stir or rock for up to 3 hours, if necessary, to facilitate dissolution. For stocks prepared in medium, vessel caps should be loose to allow for CO_2 exchange. Proceed with dosing solution preparation and dosing.
  - If a range finder experiment produces a biphasic curve, then the doses selected for the subsequent main experiments should cover the most toxic dose-response range (see **Example C1-1** – the most toxic range is 0.001 – 0.1 µg/mL) that reduces viability to 50%.

Example C1-1 Biphasic Curve



- Main Experiment (Definitive Assay)
  - Depending on the slope of the concentration-response curve estimated from the range finder experiment, the dilution/progression factor in the concentration series of the main experiment should be smaller (e.g., dilution factor of $\sqrt[6]{10} = 1.47$).
  - Cover the relevant concentration range around the IC_{50} (>0% and <100% effect) preferably with several points of a graded effect, but with a minimum of two points, one on each side of the estimated IC_{50} value, avoiding too many non-cytotoxic and/or 100%-cytotoxic concentrations.
  - Determine which test substance concentration is closest to the IC_{50} value. Use that value as a central concentration and adjust dilutions higher and lower in equal steps for the definitive assay.
  - The number of definitive tests that should be performed for a test substance is two.

6.4.4 Maximum Doses to be Tested in the Main Experiment

If minimal or no cytotoxicity was measured in the dose range finder experiment, a maximum dose for the main experiments will be established as follows:

6.4.4.1 *For test substances prepared in Chemical Dilution Medium*

- The highest test substance concentration that may be applied to the cells in the main experiments will be either 100 mg/mL, or the maximum soluble dose.
- Test substance will be weighed into a glass tube and the weight will be documented. A volume of Chemical Dilution Medium will be added to the vessel so that the concentration is 200,000 μg/mL (200 mg/mL).
- The solution is mixed using the mechanical procedures that produced solubility when performing the solubility test (See **ANNEX I**).
- If complete solubility is achieved in medium, then seven additional serial stock dosing solutions may be prepared from the 200 mg/mL 2X stock.

- If the test substance is insoluble in medium at 200 mg/mL, proceed by adding medium, in small incremental amounts, to attempt to dissolve the substance by using the sequence of mechanical procedures specified in **ANNEX I**.
- More stringent solubility procedures may be employed if needed based on results from the range finder experiment (**Section 6.4.3**). The highest soluble stock solution will be used to prepare the seven additional serial stock dosing solutions.

6.4.4.2 *For test substances prepared in either DMSO or ETOH*

- The highest test substance concentration that may be applied to the cells in the main experiments will be ≤2.5 mg/mL or less, depending upon the maximum solubility in solvent.
- Weigh the test substance into a glass tube and document the weight. Add the appropriate solvent (determined from the original solubility test) to the vessel so that the concentration is 500,000 μg/mL (500 mg/mL).
- Mix the solution using the sequence of mechanical procedures specified in **ANNEX I**.
- If complete solubility is achieved in the solvent, then seven additional serial stock dosing solutions may be prepared from the 500 mg/mL 200X stock.
- If the test substance is insoluble in solvent at 500 mg/mL, proceed by adding solvent, in small incremental amounts, to attempt to dissolve the substance by again using the sequence of mixing procedures. The highest soluble stock solution will be used to prepare the seven additional serial stock dosing solutions.

If precipitates are observed in the 2X dilutions, continue with the experiment and make the appropriate observations and documentation.

6.4.4.3 *Test Substance Dilutions*

The dosing factor of 3.16 (= $\sqrt[2]{10}$) divides a log into two equidistant steps, 2.15 (= $\sqrt[3]{10}$) into three steps, 1.78 ($\sqrt[4]{10}$) into four steps, 1.47 (= $\sqrt[6]{10}$) into six steps, and 1.21 (= $\sqrt[12]{10}$) into 12 steps.

Example C1-2 Example of Decimal Geometric Concentration Series for Factor 1.47

10						31.6						100
10				21.5				46.4				100
10		14.7		21.5		31.6		46.4		68.1		100
10	12.1	14.7	17.8	21.5	26.1	31.6	38.3	46.4	56.2	68.1	82.5	100

An example of decimal geometric concentration series for factor 1.47: Dilute 1 volume of the highest concentration by adding 0.47 volumes of diluent. After equilibration, dilute 1 volume of this solution by adding 0.47 volumes of diluent...(etc.).

6.5 Test Procedure

6.5.1 96-Well Plate Configuration

The 3T3 NRU assay for test substances will use the 96-well plate configuration as shown in **Figure C1-1**.

6.5.2 Application of Test Substance and Positive Control

6.5.2.1 *Application of Test Substance*

- Two optional methods for rapidly applying the 2X dosing solutions onto the 96-well plates may be utilized.
  - Add each of the 2X dosing solutions into labeled, sterile reservoirs (e.g., Corning/Costar model 4870 sterile polystyrene 50 mL reagent reservoirs; or Corning/Transtar model 4878 disposable reservoir liners, 8-channel; or other multichannel reservoirs).
  - Use a *dummy plate* (i.e., an empty sterile 96-well plate) prepared to hold the dosing solutions immediately prior to treatment of the test plate (with cells). The test substance and control dosing solutions

should be dispensed into the dummy plate in the same pattern/order as will be applied to the plate containing cells. More volume than needed for the test plate (i.e., greater than 50 µL/well) should be in the wells of the dummy plate.

- At the time of treatment initiation, use a multi-channel micropipettor to transfer the 2X dosing solutions from the reservoirs or dummy plate to the appropriate wells on the treatment plate (as described in step c. below). These methods will ensure that the dosing solutions can be transferred rapidly to the appropriate wells of the test plate to initiate treatment times and to minimize the range of treatment initiation times across a large number of treatment plates, and to prevent *out of order* dosing.
- Do not use a multichannel repeater pipette for dispensing test substance to the plates.

- After 24 hours ± 2 hours incubation of the cells, remove Routine Culture Medium from the cells by careful inversion of the plate (i.e., dump) over an appropriate receptacle. Gently blot the plate on a sterile paper towel so that the monolayer is minimally disrupted. Do not use automatic plate washers for this procedure nor vacuum aspiration.
- Immediately add 50 µL of fresh pre-warmed Routine Culture Medium to all of the wells, including the blanks.
- Fifty microliters (50 µL) of dosing solution will be rapidly transferred from the 8-channel reservoir (or dummy plate) to the appropriate wells of the test plate using a single delivery multi-channel pipettor. For example, the VC may be transferred first (into columns 1, 2, 11, and 12), followed by the test substance dosing solutions from lowest to highest dose, so that the same pipette tips on the multi-channel pipettor can be used for the whole plate. The Vehicle Control blank (VCb) wells (column 1, column 12, wells A2, A11, H2, H11) will receive the Vehicle Control dosing solutions (which should include any solvents used).

- Blanks for wells A3 – A10 and H3 – H10 shall receive the appropriate test substance solutions for each concentration (e.g., wells A3 and H3 receive C_1 solution).
- Incubate cells for 48 hours ± 0.5 hours (37 ºC ± 1 ºC, 90% ± 10% humidity, and 5.0% ± 1.0% CO_2/air).

6.5.2.2 *Application of Positive Control*

- For each set of test substance plates used in an assay, prepare a separate plate of positive control concentrations. A separate plate for the positive controls is used so that a complete dose response curve, rather than a single point estimate, can be obtained. This will assist with troubleshooting the experiment, if the need arises.
- If multiple sets of test substance plates are set up, clearly designate the positive control plates for each set; each set will be an individual entity.
- The Study Director will decide how many test substance plates will be run with a positive control plate. This plate will follow the same schedule and procedures as used for the test substance plates (including appropriate test substance concentrations in the appropriate wells and meeting test acceptance criteria – see **Sections 6.5.1**, **6.5.2**, **and 6.5.5**).

6.5.3 Microscopic Evaluation

- After at least 46 hours of treatment, examine each plate under a phase contrast microscope to identify systematic cell seeding errors and growth characteristics of control and treated cells. Record any changes in morphology of the cells due to the cytotoxic effects of the test substance, but do not use these records for any quantitative measure of cytotoxicity. Undesirable growth characteristics of control cells may indicate experimental error and may be cause for rejection of the assay. Substances

that may etch the plastic or *film out*[7] in medium should be identified and noted.

- Use the following Visual Observations Codes (**Table C1-2**) in the description of cell culture conditions. Numerical scoring of the cells should be determined and documented in the study workbook and in the appropriate section of the Microsoft Excel® template.

Table C1-2 Visual Observations Codes

Note Code	Note Text
1	Normal Cell Morphology
2	Low Level of Cell Toxicity
3	Moderate Level of Cell Toxicity
4	High level of Cell Toxicity
1P	Normal Cell Morphology with Precipitate
2P	Low Level of Cell Toxicity with Precipitate
3P	Moderate Level of Cell Toxicity with Precipitate
4P	High level of Cell Toxicity with Precipitate
5P	Unable to View Cells Due to Precipitate

6.5.4 Measurement of NRU

- Carefully remove (i.e., dump) the medium with test substance and rinse the cells very carefully with 250 µL pre-warmed D-PBS.
- Remove the rinsing solution by dumping and remove excess by gently blotting on paper towels.
- Add 250 µL NR medium (to all wells including the blanks) and incubate (37 ºC ± 1 ºC, 90% ± 10% humidity, and 5.0% ± 1.0% CO_2/air) for 3.0 hours ± 0.1 hour.
- Observe the cells briefly during the NR incubation (e.g., between 2 and 3 hours – Study Director's discretion) for NR crystal formation. Record observations in the study workbook. Study Director can decide to reject the experiment if excessive NR crystallization has occurred.

[7] Film out indicates that a substance comes out of solution and forms a layer over the medium and the well. It is noted that if a precipitate or if a substance films out then the concentration to which the cells are being exposed to be may not be the same as the concentration placed into the test well.

- After incubation, remove the NR medium, and carefully rinse cells with 250 µL pre-warmed D-PBS.
- Decant and blot D-PBS from the plate.
- Add 100 µL NR Desorb (ETOH/acetic acid) solution to all wells, including blanks.
- Shake microtiter plate rapidly on a microtiter plate shaker for 20 – 45 minutes to extract NR from the cells and form a homogeneous solution. Plates should be protected from light by using a cover during shaking.
- Plates should be still for at least five minutes after removal from the plate shaker (or orbital mixer). If any bubbles are observed, assure that they have been ruptured prior to reading the plate. Measure the absorption (within 60 minutes of adding NR Desorb solution) of the resulting colored solution at 540 nm ± 10 nm in a microtiter plate reader (spectrophotometer), using the blanks as a reference.

Note: A mean $OD_{540 \pm 10nm}$ of 0.031 - 0.065 for the VC blanks is a target range of ODs but not a test acceptance criterion (range = mean OD ± 2.5 standard deviations; mean = 0.048; SD = 0.007; N = 233). Save raw data in the Microsoft Excel® template.

Note: The range of linearity of the microplate reader should be confirmed, as per in-house standard operating procedures. Additionally, all equipment should be calibrated according to manufacturer's instructions.

6.5.5 Quality Check of 3T3 NRU Assay

6.5.5.1 *Quality Check for PC*

- All acceptance criteria must be met by the PC for a test to be acceptable.
  - The PC (SLS) IC_{50} must be within ± two and a half (2.5) standard deviations (SD) of the historical mean established by the Test Facility and must have an r^2 (coefficient of determination) value calculated for the Hill model fit (i.e., from PRISM® software)

≥0.85. NICEATM/ECVAM study generated the following PC data:

IC_{50} mean = 41.5 µg/mL; SD = 4.8 (n = 233)

Range for IC_{50} mean ± 2.5 SD = 29.5 µg/mL – 53.5 µg/mL

- The left and right mean of the VCs do not differ by more than 15% from the mean of all VCs.
- At least one calculated cytotoxicity value >0% and ≤50% viability and at least one calculated cytotoxicity value >50% and <100% viability must be present.

6.5.5.2 *Quality Check for Test Substances*

- All acceptance criteria must be met by the test substances for a test to be acceptable.
  - The left and right mean of the VCs do not differ by more than 15% from the mean of all VCs.
  - At least one calculated cytotoxicity value >0% and ≤50% viability and at least one calculated cytotoxicity value >50% and <100% viability must be present.

Exception

If a test has only one point between 0 and 100% **and** the smallest dilution factor (i.e., 1.21) was used **and** all other test acceptance criteria were met, then the test will be considered acceptable.

Stopping Rule for Insoluble Substances

If the most rigorous solubility procedures have been performed and the assay cannot achieve adequate toxicity to meet the test acceptance criteria after three definitive trials, then the Study Director may end all testing for that particular substance.

Note: A corrected mean $OD_{540 \pm 10nm}$ of 0.183 - 0.769 for the VCs is a target range of ODs but not a test acceptance criterion (range = mean OD ± 2.5 standard deviations; mean = 0.476; SD = 0.117; N = 233).

6.5.3.3 *Checks for Systematic Cell Seeding Errors*

- To check for systematic cell seeding errors, untreated VCs are placed both at the left side (row 2) and the right side (row 11 for the test plates) of the 96-well plate. Aberrations in the cell monolayer for the VCs may reflect a volatile and toxic test substance present in the assay. If volatility is suspected, then proceed to **Section 6.5.6**.
- Checks for cell seeding errors also may be performed by examining each plate under a phase contrast microscope to assure that cell quantity is consistent.

6.5.6 Testing Volatile Substances

Although this test method is not suitable for highly volatile substances, mildly volatile substances may be tested with some success. Volatile test substances may generate vapors from the treatment medium during the test substance treatment incubation period. These vapors may become resorbed into the treatment medium in adjacent wells, such that culture wells nearest the highest doses may become contaminated by exposure. If the test substance is particularly toxic at the doses tested, the cross contamination may be evident as a significant reduction in viability in the VC cultures (i.e., VC1) adjacent to the highest test substance doses.

If potential test substance volatility is suspected (e.g., for low density liquids) or if the initial range finder test (non-sealed plate) results show evidence of toxic effects in the control cultures (i.e., >15% difference in viability between VC1 [column 2] and VC2 [column 11]), then seal the subsequent test plates using the following procedure.

- Plates and substances will be prepared as usual according to **Sections 6.4** and **6.5**.
- Immediately after the 96-well culture plate has been treated with the suspected volatile substance (**Section 6.5.2**), apply the adhesive plate sealer (e.g., using a hand, microplate roller, etc.) directly over the culture

wells. Assure that the sealer adheres to each culture well (well tops should be dry).

- Place the 96-well plate cover over the sealed plate and incubate the plate under specified conditions (**Section 6.5.2**). Note: Do not jam the plate lid over the film to avoid deforming the sealer and causing the sealer to detach from culture wells. Loose fit of the plate lid is acceptable.
- At the end of the treatment period, the plate sealer should be carefully removed to avoid spillage. Continue with the NRU assay as per **Section 6.5.4**.

6.6 Data Analysis

- The Study Director will use good biological/scientific judgment for determining *unusable* wells that will be excluded from the data analysis and provide explanations for the removal of any data from the analysis.
- A calculation of cell viability expressed as NRU is made for each concentration of the test substance by using the mean NRU of the six replicate values (minimum of four acceptable replicate well) per test concentration (blanks will be subtracted). This value is compared with the mean NRU of all VC values. Relative cell viability is then expressed as percent of untreated VC. If achievable, the eight concentrations of each substance tested will span the range of no effect up to total inhibition of cell viability.
- Data from the microtiter plate reader should be transferred to a spreadsheet template (e.g., Microsoft Excel®) that will automatically determine cell viability, calculate IC_{50} values by linear interpolation, and perform statistical analyses (including statistical identification of outliers) (see **ANNEX II** for an example spreadsheet template).
- A Hill function analysis should be performed using statistical software (e.g., GraphPad PRISM® 3.0) and a template to calculate IC_{20}, IC_{50}, and IC_{80} values (and the associated confidence limits) for each test substance.

The Hill function is recommended because all the dose-response information, rather than a few points around the IC_{50}, can be used to calculate the data. Additionally, the slope of the curve can be assessed using the Hill function.

- Dose-responses for which the toxicity plateaus as concentration increases do not fit the Hill function well when Bottom = 0. To obtain a better model fit, unconstrain the Bottom parameter so that the model calculates the Bottom value. However, when Bottom ≠ 0, the EC_{50} reported by the Hill function ≠ 50% viability since the Hill function defines EC_{50} as the point midway between Top and Bottom. To obtain the appropriate IC_{50} when Bottom ≠ 0, use the following rearranged Hill function:

$$X = \log EC_{50} - \frac{\log\left(\frac{Top - Bottom}{Y - Bottom} - 1\right)}{HillSlope}$$

X = the logarithm of concentration at 50% response, $\log EC_{50}$ = logarithm of concentration at the response midway between Top and Bottom, Top = the maximum response, Bottom = minimum response, Y = 50 (i.e., 50% response), and HillSlope = the steepness of the curve.

Note: IC_{50} values are used in a regression formula to predict the LD_{50} value of a test substance as an estimate of the starting dose for an acute oral toxicity test.

7.0 BACKGROUND REFERENCE MATERIALS

Hackenberg, U. and H. Bartling. 1959. Messen und Rechnen im pharmakologischen Laboratorium mit einem speziellen Zahlensystem (WL24-System). Arch. Exp. Pathol. Pharmakol. 235: 437-463.

Spielmann, H., S. Gerner, S. Kalweit, R. Moog, T. Wirnserberger, K. Krauser, R. Kreiling, H. Kreuzer, N.P. Luepke, H.G. Miltenburger, N. Müller, P. Murmann, W. Pape, B. Siegmund, J. Spengler, W. Steiling, and F.J. Wiebel. 1991. Interlaboratory assessment of alternatives to the Draize eye irritation test in Germany. Toxicol. *In Vitro* 5: 539-542.

Test Method Protocol for Solubility Determination. In Vitro Cytotoxicity Validation Study. Phase III. August 29, 2003. Prepared by The National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). Available: http://iccvam.niehs.nih.gov/methods/invitro.htm [accessed 15 September 2006].

U. S. Environmental Protection Agency. 1996. Product Properties Test Guidelines. OPPTS 803.7840. Water Solubility: Column Elution Method; Shake Flask Method. EPA712-C-96-041, Prevention, Pesticides and Toxic Substances, Washington DC.

Test Method Protocol for the 3T3 Neutral Red Uptake Cytotoxicity Test. A Test for Basal Cytotoxicity for an In Vitro Validation Study. November 4, 2004. Prepared by the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). Available: http://iccvam.niehs.nih.gov/methods/invitro.htm [accessed 15 September 2006].

8.0 REFERENCES

Hartung T, Balls M, Bardouille C, Blanck O, Coecke S, Gstrauthaler G, and Lewis D. 2002. Good Cell Culture Practice: ECVAM Good Cell Culture Practice Task Force Report 1. Altern Lab Anim 30: 407-414. Available: http://ecvam.jrc.it/publication/index5007.html [accessed 02 September 2005].

ANNEX I

TEST METHOD PROCEDURE

Solubility Determination of Test Substances

1.0 PROPOSAL

This procedure was designed to identify the solvent that would provide the highest soluble concentration of a test substance so there would be uniform availability of the substance to cells used for *in vitro* basal cytotoxicity testing. The solubility exercises can be performed in a routine and repeatable manner and provide guidelines to effectively prepare test substances for toxicity testing in the Neutral Red Uptake (NRU) test methods. All individuals involved in solubility assessments should be trained so as to understand solvent and solubility issues.

2.0 TEST SYSTEM

The solubility test procedure is based on attempting to dissolve substances in various solvents with increasingly rigorous mechanical techniques. The solvents to be used, in the order of preference, are cell culture medium, dimethyl sulfoxide (DMSO), and ethanol (ETOH). Determination of whether a test substance has dissolved can be based on visual observation or through the use of a microscope. A test substance has dissolved if the solution is clear and shows no signs of cloudiness or precipitation.

3.0 PROCEDURES

Preparation of the 3T3 medium will follow all procedures in the 3T3 NRU protocol.

3.1 Materials

See Section 6.1 of Test Method Protocol for the BALB/c 3T3 NRU Cytotoxicity Test Method Protocol.

3.2 Preparation of Media and Solutions

See Section 6.2 of Test Method Protocol for the BALB/c 3T3 NRU Cytotoxicity Test Method Protocol. All solutions glassware, pipettes, etc., should be sterile and all procedures should be carried out under aseptic conditions and in the sterile environment of a laminar flow cabinet (biological hazard standard). All methods and procedures should be adequately documented.

3.3 Determination of Solubility

- Solubility should be determined in a step-wise procedure that involves attempting to dissolve a test substance at a relatively high concentration with the sequence of mechanical procedures specified in **Annex I, Section 3.5**. **Table C1-3** and **Figures C1-2** and **C1-3** illustrate the step-wise procedures.
- The hierarchy of preference of solvent for dissolving test substances is medium, DMSO, and then ETOH. If the substance does not dissolve in the solvent, the volume of solvent is increased so as to decrease the test substance concentration by a factor of 10, and then the sequence of mechanical procedures are repeated in an attempt to solubilize the substance at the lower concentrations.
- For testing solubility in medium, the starting concentration is 200,000 μg/mL (i.e., 200 mg/mL) in Tier 1, but for DMSO and ETOH the starting concentration is 200,000 μg/mL (i.e., 200 mg/mL) in Tier 3.

Table C1-3 Determination of Solubility in Chemical Dilution Medium, DMSO, or ETOH

Tier	1	2	3	4	5	6
Total Volume Chemical Dilution Medium	0.5 mL	0.5 mL	5 mL	50 mL		
Concentration of Test Substance Tier 1: Add ~ 100 mg to a tube. Add enough medium to equal Tier 1 volume. If insoluble, go to Tier 2. Tier 2: Add ~10 mg to another tube. Add enough medium to equal the first volume. Dilute to subsequent volumes if necessary.	200,000 µg/mL (200 mg/mL)	20,000 µg/mL (20 mg/mL)	2,000 µg/mL (2 mg/mL)	200 µg/mL (0.20 mg/mL)		
Total Volume DMSO/ETOH			0.5 mL	5 mL	50 mL	
Concentration of Test Substance (Add ~100 mg to a large tube. Add enough DMSO or ETOH to equal the first volume. Dilute with subsequent volumes if necessary.)			200,000 µg/mL (200 mg/mL)	20,000 µg/mL (20 mg/mL)	2,000 µg/mL (2 mg/mL)	
Total Volume DMSO/ETOH						50 mL
Concentration of Test Substance (Add ~10 mg to a large tube. Add enough DMSO or ETOH to equal 50 mL.)						200 µg/mL (0.2 mg/mL)
EQUIVALENT CONCENTRATION ON CELLS	100,000 µg/mL (100 mg/mL)	10,000 µg/mL (10 mg/mL)	1000 µg/mL (1 mg/mL)	100 µg/mL (0.1 mg/mL)	10 µg/mL (0.01 mg/mL)	1 µg/mL (0.001 mg/mL)

Abbreviations: DMSO: Dimethyl sulfoxide; ETOH: Ethanol.
Note: The amounts of test substance weighed and Chemical Dilution Medium added may be modified from the amounts given above, provided that the targeted concentrations specified for each tier are tested.

Figure C1-2 Solubility Step-Wise (Tiered) Procedure

TIER 1

STEP 1:	200 mg/mL test substance (TS) in 0.5 mL Chemical Dilution Medium • if TS soluble in medium, then **STOP**. • if TS insoluble in medium, then go to STEP 2.

TIER 2

STEP 2:	20 mg/mL TS in 0.5 mL Chemical Dilution Medium • if TS soluble, then **STOP**. • if TS insoluble, then go to STEP 3.

TIER 3

STEP 3:	200 mg/mL TS in DMSO • if TS soluble, then **STOP**. • if TS insoluble, test at 200 mg/mL in ETOH. – if TS soluble, then **STOP.** – If TS insoluble, go to STEP 4.

TIER 4

STEP 4:	0.2 mg/mL TS in medium (one or both) – increase volume from STEP 2 by 10 (i.e., to 50 mL) • if TS soluble in both media, then **STOP**. • if TS insoluble in one medium, test at 20 mg/mL in DMSO – increase volume from STEP 3 by 10 (i.e., to 5 mL). – if TS soluble, then **STOP.** – if TS insoluble, test at 20 mg/mL in ETOH – increase volume from STEP 3 by 10 (i.e., to 5 mL). ▪ if TS soluble, then **STOP.** ▪ if TS insoluble, then go to STEP 5.

TIER 5

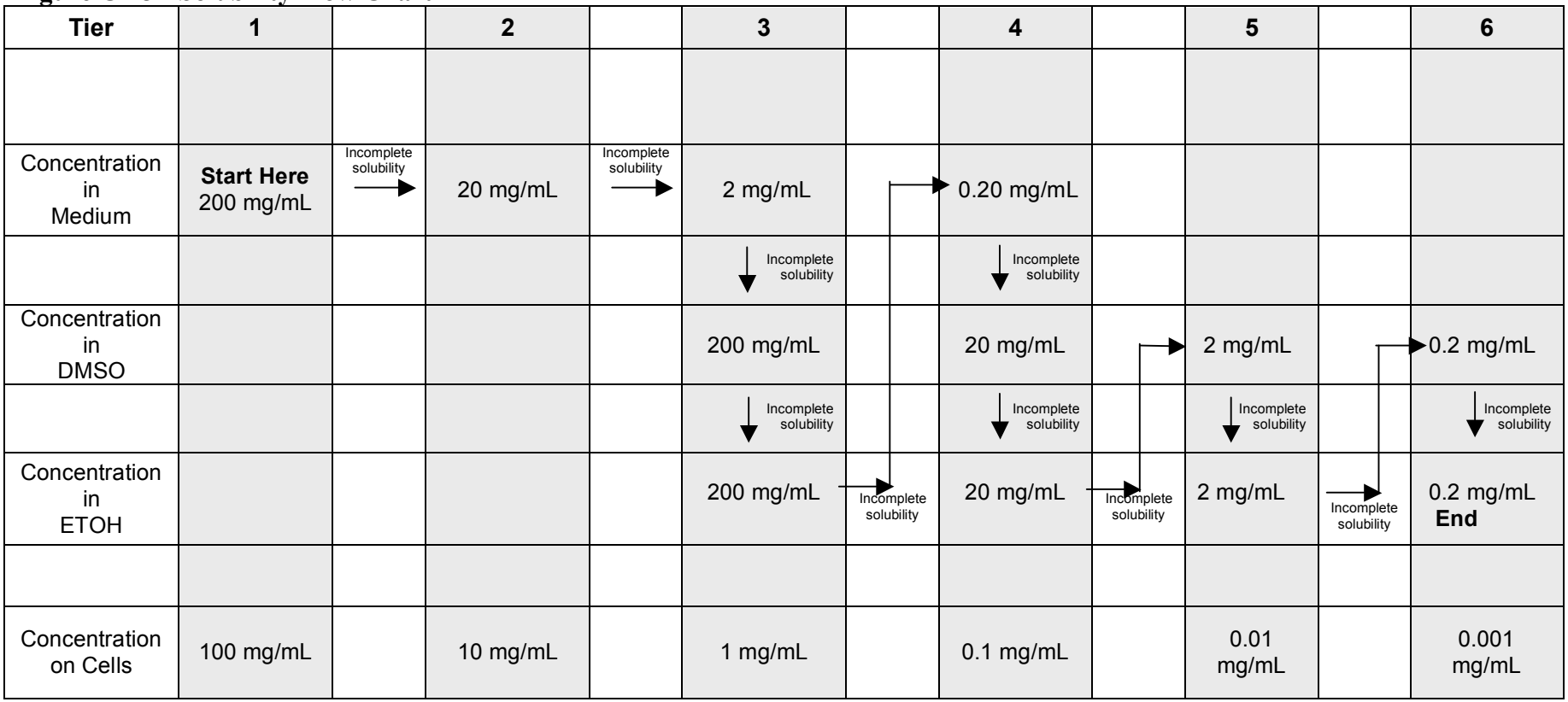
STEP 5:	2 mg/mL TS in DMSO – increase volume from STEP 4 by 10 (i.e., to 50 mL) • if TS soluble, then **STOP.** • if TS insoluble, test at 2 mg/mL in ETOH – increase volume from STEP 4 by 10 (i.e., to 50 mL). – if TS soluble, then **STOP.** – if TS insoluble, then go to STEP 6.

TIER 6

STEP 6:	0.2 mg/mL TS in 50 mL DMSO • if TS soluble, then **STOP.** • if TS insoluble, test at 0.2 mg/mL in 50 mL ETOH – **STOP**

Abbreviations: DMSO: Dimethyl sulfoxide; ETOH: Ethanol.

Figure C1-3 Solubility Flow Chart



Notes: 3T3 Medium - Dulbecco's Modification of Eagle's Medium, with supplements, for 3T3 mouse fibroblasts

3.4 Methods

3.4.1 Tier 1

- Tier 1 begins with testing 200 mg/mL in Chemical Dilution Medium (see **Table C1-3**).
  - Weigh approximately 100 mg (100,000 μg) of the test substance into a glass tube. Document the test substance weight.
  - Add approximately 0.5 mL of medium into the tube so that the concentration is 200,000 μg/mL (200 mg/mL).
  - Mix the solution as specified in **Annex I, Section 3.5**. If complete solubility is achieved, then additional solubility procedures are not needed.

3.4.2 Tier 2

- If the test substance is insoluble in Tier 1 at 200 mg/mL, then proceed to Tier 2.
  - Weigh approximately 10 mg (10,000 μg) of the test substance into a glass tube. Document the substance weight.
  - Add approximately 0.5 mL of medium into the tube so that the concentration is 20,000 μg/mL (20 mg/mL).
  - Mix the solution as specified in **Annex I, Section 3.5**. If complete solubility is achieved, then additional solubility procedures are not needed.

3.4.3 Tier 3

- If the test substance is insoluble in Chemical Dilution Medium, proceed to Tier 3.
  - Add enough medium, approximately 4.5 mL, to attempt to dissolve the substance at 2 mg/mL by using the sequence of mixing procedures. If the test substance dissolves in medium at 2 mg/mL, no further procedures are necessary.
  - If the test substance does not dissolve in medium, weigh out approximately 100 mg test substance in a second glass tube and

913 add enough DMSO to make the total volume approximately 0.5
914 mL (for 200 mg/mL) and mix the solution as specified in **Annex I,**
915 **Section 3.5**.
916 o If the test substance does not dissolve in DMSO, weigh out
917 approximately 100 mg test substance in another glass tube and add
918 enough ETOH to make the total volume approximately 0.5 mL (for
919 200 mg/mL) and mix the solution as specified in **Annex I,**
920 **Section 3.5**.
921 o If the substance is soluble in either solvent, no additional solubility
922 procedures are needed.
923 3.4.4 Tier 4
924 • If the substance is insoluble in Chemical Dilution Medium, DMSO, or
925 ETOH at Tier 3, then continue to Tier 4 in **Table C1-3**.
926 o Add enough solvent to increase the volume of the three (or four)
927 Tier 2 solutions by 10 and attempt to solubilize again using the
928 sequence of mixing procedures. If the test substance dissolves, no
929 additional solubility procedures are necessary.
930 o If the test substance does not dissolve, continue with Tier 5 and, if
931 necessary, Tier 6 using DMSO and ETOH.
932 3.4.5 Tier 5
933 • Tier 5 begins by diluting the Tier 4 samples with DMSO or ETOH to
934 bring the total volume to 50 mL. The mixing procedures are again
935 followed to attempt to solubilize the substance.
936 3.4.6 Tier 6
937 • Tier 6 is performed, if necessary, by weighing out another two samples of
938 test substance at ~10 mg each and adding ~50 mL DMSO or ETOH for a
939 200 μg/mL solution, and following the mixing procedures.
940
941 Example
942 • If complete solubility is not achieved at 20,000 μg/mL in Chemical
943 Dilution Medium at Tier 2 using the mixing procedures, then the

procedure continues to Tier 3 by diluting the solution to 5 mL with medium and mixing again.

- If the substance is not soluble in Chemical Dilution Medium, two samples of ~ 100 mg test substance are weighed to attempt to solubilize in DMSO and ETOH at 200,000 μg/mL (i.e., 200 mg/mL). Solutions are mixed following the sequence of procedures prescribed in **Annex I, Section 3.5** in an attempt to dissolve.
- If solubility is not achieved at Tier 3, then the solutions prepared in Tier 3 are diluted by 10 so as to test 200 μg/mL in media, and 20,000 μg/mL in DMSO and ETOH. This advances the procedure to Tier 4. Solutions are again mixed in an attempt to dissolve.
- If solubility is not achieved in Tier 4, the procedure continues to Tier 5, and to Tier 6, if necessary (see **Figures C1-2** and **C1-3** and **Table C1-3**).

3.5 Mechanical Procedures

The following hierarchy of mixing procedures will be followed to dissolve the test substance:

- Add test substance to solvent as in Tier 1 of **Table C1-3**. (Test substance and solvent should be at room temperature.)
- Gently mix at room temperature. Vortex the tube (1 –2 minutes).
- If test substance has not dissolved, use waterbath sonication for up to 5 minutes.
- If test substance is not dissolved after sonication, then warm solution to 37 °C for 5 - 60 minutes. This can be performed by warming tubes in a 37 °C waterbath or in a CO_2 incubator at 37 °C. The solution may be stirred during warming (stirring in a CO_2 incubator will help maintain proper pH).
- Proceed to Tier 2 (and Tiers 3-6, if necessary of **Table C1-3** and repeat procedures 2-4).

The preference of solvent for dissolving test substances is Chemical Dilution Medium, DMSO, and then ETOH. Thus, if all solvents for a particular tier are tested simultaneously and a test substance dissolves in more than one solvent, then the choice of solvent follows this hierarchy. For example, if, at any tier, a substance were soluble in Chemical Dilution Medium and DMSO, the choice of solvent would be medium. If the substance were insoluble in medium, but soluble in DMSO and ETOH, the choice of solvent would be DMSO.

[This Page Intentionally Left Blank]

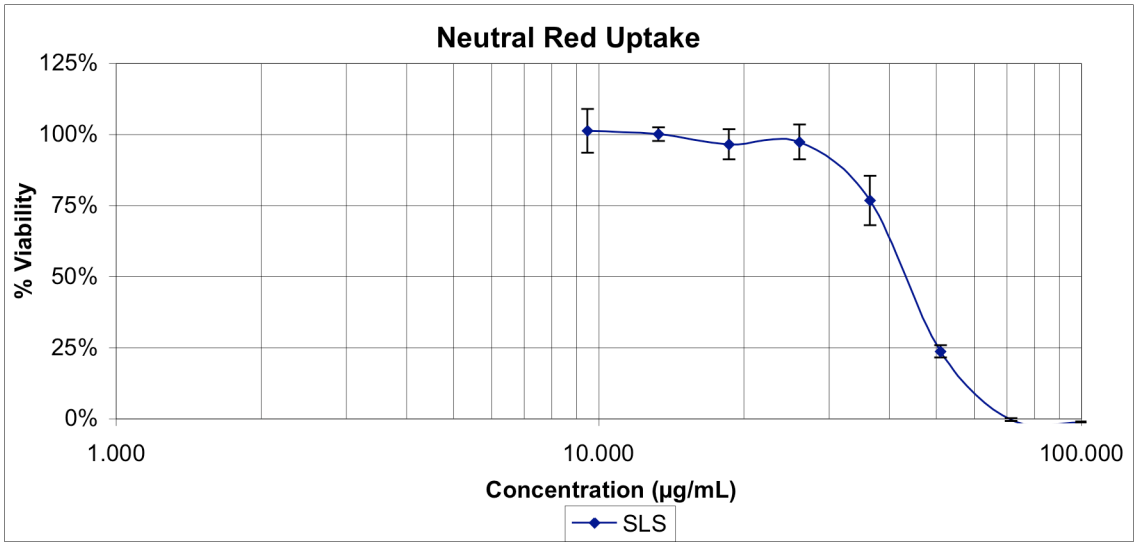
ANNEX II

Microsoft EXCEL® Example Spreadsheet Template

Test Facility :	A					Study Number.:	A1					
Chemical Code :	SLS					96-Well Plate ID :	A11					
2nd Chem. Code*:	11					Experiment ID :	XX					
					96-WELL PLATE MAP							
	1	2	3	4	5	6	7	8	9	10	11	12
A	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank
B	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
C	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
D	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
E	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
F	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
G	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
H	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank
					RAW ABSORBANCE DATA (OD_{550})							
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.044	0.044	0.045	0.045	0.045	0.046	0.051	0.057	0.057	0.043	0.041	0.044
B	0.042	0.456	0.043	0.043	0.130	0.300	0.395	0.414	0.418	0.402	0.401	0.042
C	0.043	0.407	0.042	0.041	0.130	0.294	0.383	0.382	0.413	0.375	0.385	0.044
D	0.043	0.438	0.042	0.043	0.147	0.337	0.409	0.404	0.438	0.436	0.391	0.047
E	0.044	0.448	0.041	0.045	0.132	0.321	0.429	0.414	0.416	0.420	0.441	0.042
F	0.045	0.411	0.040	0.042	0.127	0.375	0.397	0.402	0.422	0.447	0.403	0.043
G	0.041	0.405	0.043	0.040	0.124	0.361	0.444	0.442	0.425	0.448	0.405	0.044
H	0.041	0.041	0.048	0.042	0.042	0.044	0.042	0.042	0.040	0.044	0.041	0.041
Max	0.045	0.456	0.043	0.045	0.147	0.375	0.444	0.442	0.438	0.448	0.441	0.047
Min	0.041	0.405	0.040	0.040	0.124	0.294	0.383	0.382	0.413	0.375	0.385	0.041
Next Max	0.044	0.448	0.042	0.043	0.132	0.361	0.429	0.414	0.425	0.447	0.405	0.044
Next Min	0.042	0.407	0.041	0.041	0.127	0.300	0.395	0.402	0.416	0.402	0.391	0.042
Rmax	-0.250	-0.157	-0.333	-0.400	-0.652	-0.173	-0.246	-0.467	-0.520	-0.014	-0.643	-0.500
Rmin	0.250	0.039	0.333	0.200	0.130	0.074	0.197	0.333	0.120	0.370	0.107	0.167
					CORRECTED ABSORBANCE (Sample OD_{550} - Mean Blank OD_{550})							
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.001	0.001	-0.002	0.002	0.002	0.001	0.005	0.008	0.009	-0.001	-0.002	0.001
B	-0.001	0.413	-0.004	-0.001	0.087	0.255	0.349	0.365	0.370	0.359	0.358	-0.001
C	0.000	0.364	-0.005	-0.003	0.087	0.249	0.337	0.333	0.365	0.332	0.342	0.001
D	0.000	0.395	-0.005	-0.001	0.104	0.292	0.363	0.355	0.390	0.393	0.348	0.004
E	0.001	0.405	-0.006	0.002	0.089	0.276	0.383	0.365	0.368	0.377	0.398	-0.001
F	0.002	0.368	-0.007	-0.001	0.084	0.330	0.351	0.353	0.374	0.404	0.360	0.000
G	-0.002	0.362	-0.004	-0.004	0.081	0.316	0.398	0.393	0.377	0.405	0.362	0.001
H	-0.002	-0.002	0.002	-0.001	-0.001	-0.001	-0.005	-0.008	-0.009	0.001	-0.002	-0.002
Mean Blank =	0.043		0.047	0.044	0.044	0.045	0.047	0.050	0.049	0.044		
					RELATIVE VIABILITY (% OF VEHICLE CONTROL)							
	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		110.7%	-0.9%	-0.1%	23.2%	68.4%	93.4%	97.7%	99.0%	96.1%	96.0%	
C		97.6%	-1.2%	-0.7%	23.2%	66.7%	90.2%	89.1%	97.7%	88.9%	91.7%	
D		105.9%	-1.2%	-0.1%	27.7%	78.3%	97.2%	95.0%	104.4%	105.2%	93.3%	
E		108.6%	-1.5%	0.4%	23.7%	74.0%	102.5%	97.7%	98.5%	100.9%	106.7%	
F		98.7%	-1.7%	-0.4%	22.4%	88.5%	94.0%	94.5%	100.1%	108.2%	96.5%	
G		97.1%	-0.9%	-0.9%	21.6%	84.7%	106.5%	105.2%	100.9%	108.4%	97.1%	
H												

Test Facility :	A					Study Number.:	A1				
Chemical Code :	SLS					96-Well Plate ID :	A11				
2nd Chem. Code*:	11					Experiment ID :	XX				
	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	
Conc. (µg/mL) :	0.0	100	71.4	51.0	36.4	26.0	18.6	13.3	9.49	0.0	
Mean Corr. OD :	0.385	-0.005	-0.001	0.088	0.286	0.363	0.360	0.374	0.378	0.361	
SD :	0.023	0.001	0.002	0.008	0.033	0.023	0.020	0.009	0.029	0.020	
Mean Vehicle Control :	0.373										
Mean Blank :	0.043										
% of Vehicle Control :	103.1%	-1.3%	-0.3%	23.6%	76.8%	97.3%	96.5%	100.1%	101.3%	96.9%	
SD :	6.0%	0.3%	0.5%	2.1%	8.7%	6.2%	5.3%	2.4%	7.7%	5.2%	
% CV :	5.86%	-25.1%	-150%	9.09%	11.4%	6.33%	5.47%	2.39%	7.59%	5.41%	
Mean VC - VC1 (%) :	-3.1%										
Mean VC - VC2 (%) :	3.1%										
Mean Absolute OD :	0.416										
				Visual Observations							
	VC	C1	C2	C3	C4	C5	C6	C7	C8		
ENTER CODES:	1	4	4	3	2	1	1	1	1		
				Interpolated IC_{50} :		**4.32E+01**	**µg/mL**				

1000



1001

1002

TEST CHEMICAL

Test Facility :	A	Study Number.:	A1
Chemical Code :	SLS	96-Well Plate ID :	A11
2nd Chem. Code*:	11	Experiment ID :	XX

* Testing Facility Accession Code, if applicable

PREPARATION OF TEST CHEMICAL

Solvent:	Medium	Dilution factor:	1.4	
Solvent Conc. (%, v/v) in dosing solutions :	N/A	Highest Stock Conc.:	20,000	µg/mL

Aids used to dissolve : ☐ Vortexing ☐ sonication ☐ heating to 37C

pH (highest medium stock or 2X dosing solution) : 8.0

Medium Clarity/Color (highest 2X dosing solution): clear red If ppt, note lowest conc.:

Concentration Series (µg/mL)

C1	C2	C3	C4	C5	C6	C7	C8
100	71.4	51.0	36.4	26.0	18.6	13.3	9.49

Positive Control (SLS) 100 - 9.49 µg/mL

CELL LINE/TYPE

Name:	BALB/c 3T3	Supplier:	ATCC	Lot No.	not provided
Passage No.:	69	Passage No. in Assay:	75	Proliferating/frozen	24-May-02

CELL CULTURE CONDITIONS

Medium:	DMEM	Supplier:		Lot No.:	
Serum:	NCS	Supplier:		Lot No.:	
Serum Conc.:		Growth Medium:	10%	Treatment Medium:	0%

TEST ACCEPTANCE CRITERIA

No. of values >50% and <100%:	3	No. of values >0% and ≤50%:	1	Accept?	YES
VC: % Difference between Col 2 and mean VC.:			-3%	Accept?	YES
PC: Hill Function R^2 Value of SLS:		0.99		Accept?	YES
PC: IC_{50} of SLS:		43.2	µg/mL	Accept?	YES

TIMELINE

Cell Seeding Date	Dose Application Date	OD_{550} Determination Date

TEST RESULTS

VC: Mean Corrected OD_{550}: 0.373 Hill Function R^2 Value: 0.9869

log IC20 :	1.551E+00 µg/mL	log IC50 :	1.635E+00 µg/mL	log IC80 :	1.718E+00 µg/mL
IC20 :	3.56E+01 µg/mL	IC50 :	4.32E+01 µg/mL	IC80 :	5.22E+01 µg/mL

Test Chemical F.W. : 288.4

IC_{20} : 0.12331183 mM IC_{50} : 0.1496252 mM IC_{80} : 0.18113599 mM

[This Page Intentionally Left Blank]

APPENDIX C2

TEST METHOD PROTOCOL FOR THE NHK NRU CYTOTOXICITY TEST METHOD

[This Page Intentionally Left Blank]

ICCVAM Recommended Protocol for the Normal Human Keratinocyte (NHK) Neutral Red Uptake (NRU) Cytotoxicity Test - A Test for Basal Cytotoxicity

1.0 PURPOSE

This test method is used to evaluate the cytotoxicity of test substances using the Normal Human Keratinocyte (NHK) Neutral Red Uptake (NRU) *in vitro* cytotoxicity test. The data generated from the *in vitro* cytotoxicity assays are used to predict the starting doses for rodent acute oral systemic toxicity assays. This test method protocol outlines the procedures for performing the basal cytotoxicity test and is the result of the joint independent *in vitro* validation study organized by National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the European Centre for the Validation of Alternative Methods (ECVAM).

If changes or modifications are made to this protocol, the testing laboratory should prove that the results are comparable to those obtained when using the original protocol.

2.0 TEST SYSTEM

The NRU cytotoxicity assay procedure is based on the ability of viable cells to incorporate and bind neutral red (NR), a supravital dye. NR is a weak cationic dye that readily diffuses through the plasma membrane and concentrates in lysosomes where it electrostatically binds to the anionic lysosomal matrix. Toxicants can alter the cell surface or the lysosomal membrane to cause lysosomal fragility and other adverse changes that gradually become irreversible. Thus, cell death and/or inhibition of cell growth decreases the amount of neutral red retained by the culture. Healthy proliferating mammalian cells, when properly maintained in culture, continuously divide and multiply over time. A toxic chemical, regardless of site or mechanism of action, will interfere with this process and result in a reduction of the growth rate as reflected by cell number.

Cytotoxicity is expressed as a concentration dependent reduction of the uptake of NR after chemical exposure, thus providing a sensitive, integrated signal of both cell integrity and growth inhibition.

3.0 KEY PERSONNEL

3.1 Laboratory

- Study Director (only recommended if testing is performed under Good Laboratory Procedures [GLP])
- Laboratory Technician(s)

3.2 Testing Facility

- Scientific Advisor
- Quality Assurance Director (only necessary if testing is performed under GLP)
- Safety Manager
- Facility Management

4.0 DEFINITIONS

Hill function: a four parameter logistic mathematical model relating the concentration of test substance to the response being measured in a sigmoidal shape

$$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(\log EC_{50} - X)\text{HillSlope}}}$$

where Y = response, X = the logarithm of dose (or concentration), Bottom = the minimum response, Top the maximum response, $\log EC_{50}$ = logarithm of X at the response midway between Top and Bottom, and HillSlope = the

steepness of the curve. When Top = 100 and Bottom = 0, the EC_{50} is the concentration at 50% viability (i.e., the IC_{50})

Documentation: all methods and procedures will be noted in a study workbook; logs will be maintained for general laboratory procedures and equipment (e.g., media preparation, test substance preparation, incubator function); all optical density data obtained from the spectrophotometer plate reader will be saved in electronic and paper formats; all calculations of ICx values and other derived data will be in electronic and paper format; all data will be archived

IC_{50}: test substance concentration producing 50% inhibition of the endpoint measured (i.e., cell viability)

5.0 IDENTIFICATION OF CONTROL SUBSTANCES

5.1 Positive Control (PC)

- Sodium Lauryl Sulfate (SLS)

5.2 Vehicle Control (VC)

- Keratinocyte assay medium

5.3 Solvent Control

- VC control with solvent (i.e., keratinocyte assay medium, dimethyl sulfoxide [DMSO], or ethanol [ETOH]). DMSO is the preferred solvent for substances that are not water (i.e., assay medium) soluble.

6.0 PROCEDURES

6.1 Materials

6.1.1 Cell Line

- Normal Human Epidermal Keratinocytes (NHK) cells. Non-transformed cells; from cryopreserved primary or secondary cells (e.g., Clonetics #CC-2507 or equivalent - Cambrex [Cambrex Bio Science, 8830 Biggs Ford Road, Walkersville, MD]. Cells will be Clonetics NHK cells.[8]

6.1.2 Technical Equipment[9]

- Incubator: 37 ºC ± 1 ºC, 90% ± 10% humidity, 5.0% ± 1.0% CO_2/air
- Laminar flow clean bench/cabinet (standard: "biological hazard")
- Waterbath: 37 ºC ± 1 ºC
- Inverse phase contrast microscope
- Sterile glass tubes with caps (e.g., 5 mL)
- Centrifuge
- Laboratory balance
- 96-well plate spectrophotometer (i.e., plate reader) equipped with 540 nm ± 10 nm filter with maximum absorbance of 3
- Shaker for microtiter plates
- Cell counter or hemocytometer
- Pipetting aid
- Pipettes, pipettors (multi-channel and single channel; multichannel repeater pipette), dilution block
- Cryotubes
- Tissue culture flasks (e.g., 75 - 80 cm^2, 25 cm^2)

[8] Keratinocytes should be procured only through commercial sources and not by preparing primary culture from donated tissues.
[9] Suggested brand names/vendors are listed in parentheses. Equivalents may be used.

- 96-well flat bottom tissue culture microtiter plates (e.g., Nunc # 167 008; Corning/COSTAR tissue culture-treated)
- pH paper (wide and narrow range)
- Multichannel reagent reservoir
- Waterbath sonicator
- Magnetic stirrer
- Antistatic bar ionizer/antistatic gun (optional for neutralizing static on 96-well plates)
- Dry heat block (optional)
- Adhesive film plate sealers (e.g., Excel Scientific SealPlate™, Cat # STR-SEAL-PLT or equivalent)
- Vortex mixer
- Filters/filtration devices

Note: Prescreen tissue culture flasks and microtiter plates to ensure that they adequately support the growth of NHK cells. Use multi-channel repeater pipettes for plating cells in the 96-well plates, dispensing plate rinse solutions, NR medium, and desorb solution. Do not use the repeater pipette for dispensing test substances to the cells.

6.1.3 Chemicals, Media, and Sera

- Keratinocyte Basal Medium without Ca^{++} (e.g., KBM®, Clonetics CC-3104) that is completed by adding supplements (e.g., KBM® SingleQuots®, Clonetics CC-4131) to achieve the proper concentrations of epidermal growth factor, insulin, hydrocortisone, antimicrobial agents, bovine pituitary extract, and calcium (e.g., Clonetics Calcium SingleQuots®, 300 mM $CaCl_2$, Clonetics CC-4202).
- HEPES Buffered Saline Solution (HEPES-BSS) (e.g., Clonetics # CC-5022)
- 0.025 % Trypsin/EDTA solution (e.g., Clonetics # CC-5012)
- Trypsin Neutralizing Solution (TNS) (e.g., Clonetics # CC-5002)
- Phosphate Buffered Saline (PBS)

- Dulbecco's Phosphate Buffered Saline (D-PBS) (formulation containing calcium and magnesium cations; glucose optional)
- Neutral Red (NR) Dye – tissue culture-grade; liquid form (e.g., SIGMA N 2889); powder form (e.g., SIGMA N 4638)
- DMSO, U.S.P analytical grade (Store under nitrogen @ -20ºC)
- ETOH, U.S.P. analytical grade (100 %, non-denatured for test substance preparation; 95 % can be used for the desorb solution)
- Glacial acetic acid, analytical grade
- Hanks' Balanced Salt Solution without Ca^{2+} or Mg^{2+} (CMF-HBSS) (e.g., Invitrogen # 14170)
- Distilled H_2O or any purified water suitable for cell culture and NR desorb solution (sterile)
- Sterile/non-sterile paper towels (for blotting 96-well plates)

6.2 Preparation of Media and Solutions

Note: All solutions (except NR stock solution, NR medium and NR desorb), glassware, pipettes, etc., shall be sterile and all procedures should be carried out under aseptic conditions and in the sterile environment of a laminar flow cabinet (biological hazard standard). All methods and procedures will be adequately documented.

6.2.1 Media

Note: This protocol is based on the use of Clonetics KBM® medium and supplements. Other media may be acceptable if proper cell growth conditions can be maintained as per this protocol. Prequalify candidate media by using the keratinocyte medium prequalification in **ANNEX I**.

Routine Culture Medium/Treatment Medium: KBM® (Clonetics CC-3104) supplemented with KBM® SingleQuots® (Clonetics CC-4131) and Clonetics Calcium

SingleQuots® (CC-4202) to make 500 mL medium. Final concentrations of supplements in medium are:

- 0.0001 ng/mL Human recombinant epidermal growth factor
- 5 μg/mL Insulin
- 0.5 μg/mL Hydrocortisone
- 30 μg/mL Gentamicin
- 15 ng/mL Amphotericin B
- 0.10 mM Calcium
- 30 μg/mL Bovine pituitary extract

Complete media formulations should be kept at 2-8 °C and stored for no longer than two weeks.

KBM® SingleQuots® contain the following stock concentrations and volumes:

• 0.1 ng/mL	hEGF	0.5 mL
• 5.0 mg/mL	Insulin	0.5 mL
• 0.5 mg/mL	Hydrocortisone	0.5 mL
• 30 mg/mL	Gentamicin, 15 μg/mL Amphotericin-B	0.5 mL
• 7.5 mg/mL	Bovine Pituitary Extract (BPE)	2.0 mL

Clonetics Calcium SingleQuots® are 2 mL of 300 mM calcium.

165 μL of solution per 500 mL calcium-free medium equals 0.10 mM calcium in the medium.

6.2.2 NR Stock Solution

- The liquid tissue culture-grade stock NR Solution is the first choice (e.g., SIGMA #N2889, 3.3 mg/mL). Store liquid tissue culture-grade NR Stock

Solution at the storage conditions and shelf-life period recommended by the manufacturer.

- A stock solution can be made with powder NR dye and water (e.g., 0.33 g NR Dye powder in 100 mL H_2O) if the liquid stock form is not available. The stock should be stored in the dark at room temperature for up to two months.

6.2.3 NR Medium

EXAMPLE:

1.0 mL (3.3 mg NR dye/mL)	NR Stock Solution
99.0 mL	Routine Culture Medium (pre-warmed to 37°C)

The final concentration of the NR Medium is 33 μg NR dye/mL and aliquots will be prepared on the day of application.

Note: Filter the NR Medium (e.g., Millipore filtering, 0.2 – 0.45 μm pore size) to reduce NR crystals. Maintain aliquots of the NR Medium at 37 °C (e.g., in a waterbath) before adding to the cells and use within 60 minutes of preparation and within 15 minutes after removing from 37 °C storage. Examine the solution for crystals prior to use.

6.2.4 ETOH/Acetic Acid Solution (NR Desorb)

- 1% Glacial acetic acid solution
- 50% ETOH
- 49% H_2O

6.3 Methods

6.3.1 Cell Maintenance and Culture Procedures

- NHK cells are routinely grown as a monolayer in tissue culture grade flasks (e.g., 25 cm^2) at 37 ºC ± 1 ºC, 90% ± 10% humidity, and 5.0% ± 1.0% CO_2/air.
- Examine the cells on a daily (i.e., on workdays) basis under a phase contrast microscope, and note any changes in morphology or their adhesive properties in a study workbook. Cells should not reach confluence.
- All cell culture studies should follow good cell culture practices (Hartung et al. 2002).

6.3.2 Receipt of Cryopreserved Keratinocyte Cells

Upon receipt of cryopreserved keratinocytes, store the vial(s) of cells in a liquid nitrogen freezer until needed.

6.3.3 Thawing Cells

- Thaw cells by putting ampules into a waterbath at 37 °C ± 1 ºC. Leave for as brief a time as possible. Do not thaw cells at room temperature or by hand. Seed the thawed cells into culture flasks as quickly as possible and with minimal handling.
- Slowly (taking approximately 1-2 minutes) add 9 mL of pre-warmed Routine Culture Medium to the cells suspended in the cryoprotective solution and transfer cells into flasks containing pre-warmed Routine Culture Medium.
- Incubate at 37 ºC ± 1 ºC, 90% ± 10% humidity, and 5.0% ± 1.0% CO_2/air.
- When the cells have attached to the bottom of the flask (within 4 to 24 hours), the Routine Culture Medium should be removed and replaced with fresh Routine Culture Medium.

- Unless otherwise specified, the cells should be incubated at 37 ºC ± 1 ºC, 90% ± 10% humidity, 5.0% ± 1.0% CO_2/air and fed every 2-3 days until they exceed 50 % confluence (but less than 80 % confluent).

6.3.4 Subculture of NHK Cells to 96-Well Plates

Note: It is important that cells have overcome the lag growth phase when they are used for the test. Keratinocytes will be passaged only into the 96-well plates and will not be subcultured into flasks for use in later assays.

- When the keratinocyte culture in a 25 cm^2 flask >50% confluence (but <80% confluent; cell should not be 100% confluent), remove the medium and rinse the culture twice with 5 mL HEPES-BSS. The first rinse may be left on the cells for up to 5 minutes and the second rinse should remain on the cells for approximately 5 minutes. Discard the washing solutions.
- Add 2 mL trypsin/EDTA solution to each flask and remove after 15 to 30 seconds. Incubate the flask at room temperature for 3 to 7 minutes. When more than 50% of the cells become dislodged, rap the flask sharply against the palm of the hand.
- When most of the cells have become detached from the surface, rinse the flask with 5 mL of room temperature TNS. If more than one flask is subcultured, the same 5 mL of TNS may be used to rinse a total of up to two flasks.
- Then rinse the flask with 5 mL CMF-HBSS and transfer the cell suspension to a centrifuge tube.
- Pellet the cells by centrifugation for 5 minutes at approximately 220 x *g*. Remove the supernatant by aspiration.
- Resuspend the keratinocyte pellet by gentle trituration (to have single cells) in Routine Culture Medium. It is important to obtain a single cell suspension for exact counting. Count a sample of the cell suspension using a hemocytometer or cell counter.

- Prepare a cell suspension ~1.6 – 2.0 x10^4 cells/mL in Routine Culture Medium. Using a multi-channel pipette, dispense 125 µL Routine Culture Medium only into the peripheral wells (blanks) of a 96-well tissue culture microtiter plate. In the remaining wells, dispense 125 µL of the cell suspension (2x10^3 – 2.5x10^3 cells/well). Prepare one plate per substance to be tested (see **Figure C2-1**).
- Incubate cells (37 ºC ± 1 ºC, 90% ± 10% humidity, and 5% ± 1% CO_2/air) so that cells form a 20+% monolayer (~48-72 hours). This incubation period assures cell recovery and adherence and progression to exponential growth phase.
- Examine each plate under a phase contrast microscope to assure that cell growth is relatively even across the microtiter plate. This check is performed to identify experimental and systemic cell seeding errors. Record observations in the study workbook.

Figure C2-1 96-Well Plate Configuration for Positive Control (PC) and Test Substance Assays

	1	2	3	4	5	6	7	8	9	10	11	12
A	VCb	VCb	C_1b	C_2b	C_3b	C_4b	C_5b	C_6b	C_7b	C_8b	VCb	VCb
B	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
C	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
D	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
E	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
F	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
G	VCb	VC1	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	VC2	VCb
H	VCb	VCb	C_1b	C_2b	C_3b	C_4b	C_5b	C_6b	C_7b	C_8b	VCb	VCb

VC1 and VC2 = VEHICLE CONTROL
C_1 – C_8 = Test Substances or PC (SLS) at eight concentrations (C_1 = highest, C_8 = lowest)
C_xb = BLANKS (Test substance or PC, but contain **no** cells)
VCb = VEHICLE CONTROL BLANK (contain **no** cells)

6.3.5 Determination of Doubling Time

- Establish cells in culture and trypsinize cells as per **Section 6.3.4** for subculture. Resuspend cells in appropriate culture medium. Use **Table C2-1** to determine seeding densities.
- Seed five sets of cell culture vessels in triplicate for each cell type (e.g., 15 tissue culture dishes [60mm x 15mm]). Use appropriate volume of culture medium for the culture vessels. Note number of cells placed into each culture dish. Place dishes into the incubators (37 ºC ± 1 ºC, 90% ± 10% humidity, 5.0% ± 1.0% CO_2/air).
- After 4-6 hours (use the same initial measurement time for each subsequent doubling time experiment), remove three culture dishes and trypsinize cells.
- Count cells using a cell counter or hemocytometer. Cell viability may be determined by dye exclusion (e.g., Trypan Blue; Nigrosin). Determine the total number of cells and document.
- Repeat sampling at 24 hours, 48 hours, 72 hours, and 96 hours post inoculation. Change culture medium at 72 hours or sooner in remaining dishes if indicated by pH drop (i.e., pH <7).
- Plot cell concentration (per mL of medium) on a log scale against time on a linear scale. Determine lag time and population doubling time. The doubling time will be in the log (exponential) phase of the growth curve. Additional dishes and time are needed if the entire growth curve is to be determined (lag phase, log phase, plateau phase).

Table C2-1 Guidelines for Establishing Cell Cultures[1]

Cells/25 cm^2 flask (in approximately 5 mL) 1 flask each cell concentration	6.25×10^4 ($2500/cm^2$)	1.25×10^5 ($5000/cm^2$)	2.25×10^5 ($9000/cm^2$)
Approximate Time to Subculture	96+ hours	72 – 96 hours	48 - 72 hours
Cells to 96-Well Plates	6 – 8 plates	6 – 8 plates	6 – 8 plates

[1]Cell growth guidelines – actual growth of individual cell lots may vary.

6.4 Preparation of Test Substances

Note: Preparation under red or yellow light is recommended to preserve substances that degrade upon exposure to light.

Test substance solubility should be determined by following the procedures outlined in **ANNEX II** of this protocol.

6.4.1 Test Substances in Solution

- Equilibrate test substances to room temperature before dissolving and diluting.
- Prepare test substance immediately prior to use rather than preparing in bulk for use in subsequent tests. Ideally, the solutions must not be cloudy nor have noticeable precipitate. Each stock dilution should have at least 1-2 mL total volume to ensure adequate solution for the test wells in a single 96-well plate. The Study Director may store an aliquot (e.g., 1 mL) of the highest 2X stock solution (e.g., low solubility substances) in a freezer (e.g., -70 °C) for use in future substance analyses.
- For substances dissolved in DMSO or ETOH, the final DMSO or ETOH concentration for application to the cells must be 0.5% (v/v) in the vehicle controls and in all of the eight test concentrations. The concentration of DMSO or ETOH should be at the lowest possible concentration needed to dissolve the test substance.
- The stock solution for each test substance should be prepared at the highest concentration found to be soluble in the solubility test conducted per **ANNEX II**. Thus, the highest test concentration applied to the cells in each range finding experiment is:
    - 0.5 times the highest concentration found to be soluble in the solubility test, if the substance was soluble in medium, or
    - 1/200 the highest concentration found to be soluble in the solubility test if the substance was soluble in ETOH or DMSO

Example: Preparation of Test Substance for Range Finding Experiments Solvent Using a Log Dilution Scheme

If DMSO is determined to be the preferred solvent at Tier 3 of the solubility test (i.e., 200,000 μg/mL), dissolve the substance in DMSO at 200,000 μg/mL for the chemical stock solution. The seven lower concentrations in the range finding experiment are prepared by successive dilutions that decrease by one log unit each.

- Label eight tubes 1 – 8. Add 0.9 mL solvent (e.g., DMSO) to tubes 2 - 8.
- Prepare stock solution of 200,000 μg test substance/mL solvent in tube # 1.
- Add 0.1 mL of 200,000 μg/mL dilution from tube #1 to tube #2 to make a 1:10 dilution in solvent (i.e., 20,000 μg/mL).
- Add 0.1 mL of 20,000 μg/mL dilution from tube #2 to tube #3 to make another 1:10 dilution (i.e., 1:100 dilution from stock solution) in solvent (i.e., 2,000 μg/mL). Continue making serial 1:10 dilutions in the prepared solvent tubes.
- Since each concentration is 200 fold greater than the concentration to be tested, make a 1:100 dilution by diluting 1 part dissolved substance in each tube with 99 parts of culture medium (e.g., 0.1 mL of test substance in DMSO + 9.9 mL culture medium) to derive the eight 2X concentrations for application to NHK cells. Each 2X test substance concentration will then contain 1% (v/v) solvent.
    - The NHK cells will have 0.125 mL of culture medium in the wells prior to application of the test substance. By adding 0.125 mL of the appropriate 2X test substance concentration to the appropriate wells, the test substance will be diluted appropriately (e.g., highest concentration in well will be 1,000 μg/mL) in a total of 0.250 mL and the solvent concentration in the wells will be 0.5% (v/v).
- A test substance prepared in DMSO or ETOH may precipitate upon transfer into the Routine Culture Medium. The 2X dosing solutions should be evaluated for precipitates and the results recorded in the study workbook. It is permissible to test all of the dosing solutions in the dose

range finding assay and main experiments. However, doses containing test substance precipitates should be avoided because it creates doubt about the concentration of test substance exposed to the cells.

Document all test substance preparations in the study workbook.

6.4.2 pH of Test Substance Solutions

- Prior to or immediately after application of the test substance to the 96-well plate, measure the pH of the highest 2X dosing concentration of the test substance (i.e., C1 in the test plate, see **Figure C2-1**) in culture medium.
- Use pH paper (e.g., pH 0 - 14 to estimate and pH 5 – 10 to determine more precise value; or Study Director's discretion) for measurements. The pH paper should be in contact with the solution for approximately one minute.
- Document the pH and note the color of the 2X concentration medium (i.e., in the Microsoft Excel® template; see **ANNEX III** for an example template). Medium color for all dosing dilutions should be noted in the study workbook. Do not adjust the pH.

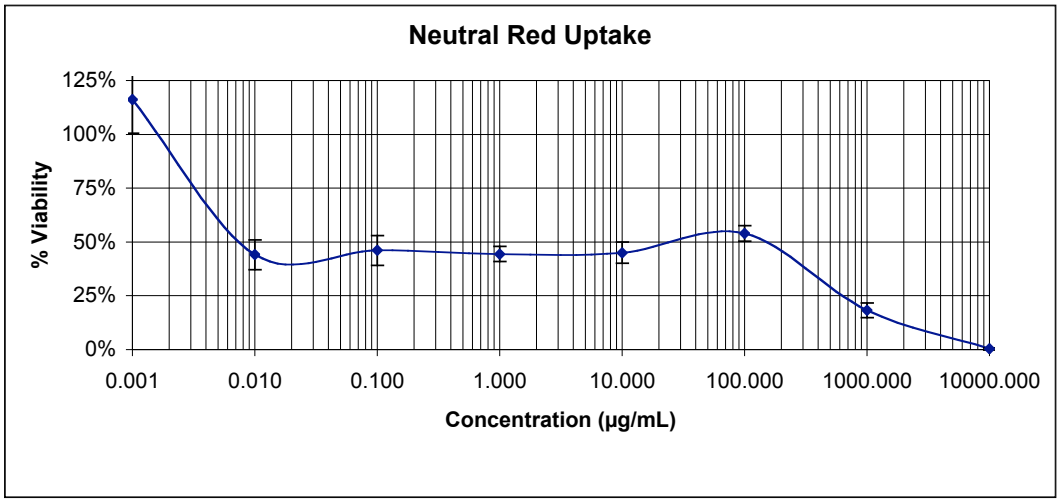
6.4.3 Concentrations of Test Substance

- Range Finder Experiment
  - Test eight concentrations of the test substance by diluting the stock solution using log dilutions (e.g., 1:10, 1:100, 1:1000, etc.).
  - If a range finder experiment does not generate enough cytotoxicity, then higher doses should be attempted. If cytotoxicity is limited by solubility, then more stringent solubility procedures to increase the stock concentration (to the maximum concentration specified in **Section 6.4.4**) should be employed.
  - Place the test substance concentration into an incubator (37 ºC ± 1 ºC, 90% ± 10% humidity, 5.0% ± 1.0% CO_2/air) and stir or rock for up to 3 hours, if necessary, to facilitate dissolution. For stocks

prepared in medium, vessel caps should be loose to allow for CO_2 exchange. Proceed with dosing solution preparation and dosing.

- If a range finder experiment produces a biphasic curve, then the doses selected for the subsequent main experiments should cover the most toxic dose-response range (see **Example C2-1** – the most toxic range is 0.001 – 0.1 µg/mL) that reduces viability to 50%.

Example C2-1 Biphasic Curve



- Main Experiment (Definitive Assay)
  - Depending on the slope of the concentration-response curve estimated from the range finder experiment, the dilution/progression factor in the concentration series of the main experiment should be smaller (e.g., dilution factor of $\sqrt[6]{10} = 1.47$).
  - Cover the relevant concentration range around the IC_{50} (>0% and <100% effect) preferably with several points of a graded effect, but with a minimum of two points, one on each side of the estimated IC_{50} value, avoiding too many non-cytotoxic and/or 100%-cytotoxic concentrations.
  - Determine which test substance concentration is closest to the IC_{50} value. Use that value as a central concentration and adjust dilutions higher and lower in equal steps for the definitive assay.

- The number of definitive tests that should be performed for a test substance is two.

6.4.4 Maximum Doses to be Tested in the Main Experiment

If minimal or no cytotoxicity was measured in the dose range finder experiment, a maximum dose for the main experiments will be established as follows:

6.4.4.1 *For test substances prepared in Routine Culture Medium*

- The highest test substance concentration that may be applied to the cells in the main experiments will be either 100 mg/mL, or the maximum soluble dose.
- Test substance will be weighed into a glass tube and the weight will be documented. A volume of Routine Culture Medium will be added to the vessel so that the concentration is 200,000 μg/mL (200 mg/mL).
- The solution is mixed using the mechanical procedures that produced solubility when performing the solubility test (See **ANNEX II**).
- If complete solubility is achieved in medium, then seven additional serial stock dosing solutions may be prepared from the 200 mg/mL 2X stock.
- If the test substance is insoluble in medium at 200 mg/mL, proceed by adding medium, in small incremental amounts, to attempt to dissolve the substance by using the sequence of mechanical procedures specified in **ANNEX II**.
- More stringent solubility procedures may be employed if needed based on results from the range finder experiment (**Section 6.4.3**). The highest soluble stock solution will be used to prepare the seven additional serial stock dosing solutions.

6.4.4.2 *For test substances prepared in either DMSO or ETOH*

- The highest test substance concentration that may be applied to the cells in the main experiments will be ≤2.5 mg/mL or less, depending upon the maximum solubility in solvent.

- Weigh the test substance into a glass tube and document the weight. Add the appropriate solvent (determined from the original solubility test) to the vessel so that the concentration is 500,000 μg/mL (500 mg/mL).
- Mix the solution using the sequence of mechanical procedures specified in **ANNEX II**.
- If complete solubility is achieved in the solvent, then seven additional serial stock dosing solutions may be prepared from the 500 mg/mL 200X stock.
- If the test substance is insoluble in solvent at 500 mg/mL, proceed by adding solvent, in small incremental amounts, to attempt to dissolve the substance by again using the sequence of mixing procedures. The highest soluble stock solution will be used to prepare the seven additional serial stock dosing solutions.

If precipitates are observed in the 2X dilutions, continue with the experiment and make the appropriate observations and documentation.

6.4.4.3 *Test Substance Dilutions*

The dosing factor of 3.16 (= $^{2}\sqrt{10}$) divides a log into two equidistant steps, 2.15 (= $^{3}\sqrt{10}$) into three steps, 1.78 ($^{4}\sqrt{10}$) into four steps, 1.47 (= $^{6}\sqrt{10}$) into six steps, and 1.21 (= $^{12}\sqrt{10}$) into 12 steps.

Example C2-2 Example of Decimal Geometric Concentration Series for Factor 1.47

10						31.6						100
10				21.5				46.4				100
10		14.7		21.5		31.6		46.4		68.1		100
10	12.1	14.7	17.8	21.5	26.1	31.6	38.3	46.4	56.2	68.1	82.5	100

An example of decimal geometric concentration series for factor 1.47: Dilute 1 volume of the highest concentration by adding 0.47 volumes of diluent. After equilibration, dilute 1 volume of this solution by adding 0.47 volumes of diluent...(etc.).

6.5 Test Procedure

6.5.1 96-Well Plate Configuration

The NHK NRU assay for test substances will use the 96-well plate configuration shown in **Figure C2-1**.

6.5.2 Application of Test Substance

6.5.2.1 *Application of Test Substance*

- Two optional methods for rapidly applying the 2X dosing solutions onto the 96-well plates may be utilized.
  - Add each of the 2X dosing solutions into labeled, sterile reservoirs (e.g., Corning/Costar model 4870 sterile polystyrene 50 mL reagent reservoirs; or Corning/Transtar model 4878 disposable reservoir liners, 8-channel; or other multichannel reservoirs).
  - Use a *dummy plate* (i.e., an empty sterile 96-well plate) prepared to hold the dosing solutions immediately prior to treatment of the test plate (with cells). The test substance and control dosing solutions should be dispensed into the dummy plate in the same pattern/order as will be applied to the plate containing cells. More volume than needed for the test plate (i.e., greater than 125 µL/well) should be in the wells of the dummy plate.
  - At the time of treatment initiation, use a multi-channel micropipettor to transfer the 2X dosing solutions from the reservoirs or dummy plate to the appropriate wells on the treatment plate (as described below). These methods will ensure that the dosing solutions can be transferred rapidly to the appropriate wells of the test plate to initiate treatment times and to minimize the range of treatment initiation times across a large number of treatment plates, and to prevent *out of order* dosing.
  - Do not use a multichannel repeater pipette for dispensing test substance to the plates.

- After 48 - 72 hours (i.e., after cells attain 20+ % confluency [see **Section 6.3.4**]) incubation of the cells, add 125 µl of the appropriate concentration of test substance, the PC, or the VC (see **Figure C2-1** for the plate configuration) directly to the test wells. Do not remove Routine Culture Medium for re-feeding the cells.
- The dosing solutions will be rapidly transferred from the 8-channel reservoir (or dummy plate) to the test plate using a single delivery multi-channel pipettor. For example, the VC may be transferred first (into columns 1, 2, 11, and 12), followed by the test substance dosing solutions from lowest to highest dose, so that the same pipette tips on the multi-channel pipettor can be used for the whole plate. The Vehicle Control blank (VCb) wells (column 1, column 12, wells A2, A11, H2, H11) will receive the Vehicle Control dosing solutions (which should include any solvents used).
- Blanks for wells A3 – A10 and H3 – H10 shall receive the appropriate test substance solution for each concentration (e.g., wells A3 and H3 receive C_1 solution).
- Incubate cells for 48 hours ± 0.5 hours (37 ºC ± 1 ºC, 90% ± 10% humidity, and 5.0% ± 1.0% CO_2/air).

6.5.2.2 *Application of Positive Control*

- For each set of test substance plates used in an assay, prepare a separate plate of positive control concentrations. A separate plate for the positive controls is proposed so that a complete dose response curve, rather than a single point estimate, can be obtained. This will assist with troubleshooting, if the need arises.
- If multiple sets of test substance plates are set up, clearly designate the positive control plates for each set; each set will be an individual entity.
- The Study Director will decide how many test substance plates will be run with a positive control plate. This plate will follow the same schedule and procedures as used for the test substance plates (including appropriate test

substance concentrations in the appropriate wells and meeting test acceptance criteria – see **Sections 6.5.1**, **6.5.2**, **and 6.5.5**).

6.5.3 Microscopic Evaluation

- After at least 46 hours of treatment, examine each plate under a phase contrast microscope to identify systematic cell seeding errors and growth characteristics of control and treated cells. Record any changes in morphology of the cells due to the cytotoxic effects of the test substance, but do not use these records for any quantitative measure of cytotoxicity. Undesirable growth characteristics of control cells may indicate experimental error and may be cause for rejection of the assay. Substances that may etch the plastic or *film out*[10] in medium should be identified and noted.
- Use the following Visual Observations Codes (**Table C2-2**) in the description of cell culture conditions. Numerical scoring of the cells should be determined and documented in the study workbook and in the appropriate section of the Microsoft Excel® template.

Table C2-2 Visual Observations Codes

Note Code	Note Text
1	Normal Cell Morphology
2	Low Level of Cell Toxicity
3	Moderate Level of Cell Toxicity
4	High level of Cell Toxicity
1P	Normal Cell Morphology with Precipitate
2P	Low Level of Cell Toxicity with Precipitate
3P	Moderate Level of Cell Toxicity with Precipitate
4P	High level of Cell Toxicity with Precipitate
5P	Unable to View Cells Due to Precipitate

[10] Film out indicates that a substance comes out of solution and forms a layer over the medium and the well. It is noted that if a precipitate or if a substance *films out* then the concentration to which the cells are being exposed to be may not be the same as the concentration placed into the test well.

6.5.4 Measurement of NRU

- Carefully remove (i.e., dump) the Routine Culture Medium with test substance and rinse the cells very carefully with 250 µL pre-warmed D-PBS.
- Remove the rinsing solution by dumping and remove excess by gently blotting on paper towels.
- Add 250 µL NR medium (to all wells including the blanks) and incubate (37 ºC ± 1 ºC, 90% ± 10% humidity, and 5.0% ± 1.0% CO_2/air) for 3 hours ± 0.1 hour.
- Observe the cells briefly during the NR incubation (e.g., between 2 and 3 hours – Study Director's discretion) for NR crystal formation. Record observations in the study workbook. Study Director can decide to reject the experiment if excessive NR crystallization has occurred.
- After incubation, remove the NR medium, and carefully rinse cells with 250 µL pre-warmed D-PBS.
- Decant and blot D-PBS from the plate.
- Add exactly 100 µL NR Desorb (ETOH/acetic acid) solution to all wells, including blanks.
- Shake microtiter plate rapidly on a microtiter plate shaker for 20 – 45 minutes to extract NR from the cells and form a homogeneous solution. Plates should be protected from light by using a cover during shaking.
- Plates should be still for at least five minutes after removal from the plate shaker (or orbital mixer). If any bubbles are observed, assure that they have been ruptured prior to reading the plate. Measure the absorption (within 60 minutes of adding NR Desorb solution) of the resulting colored solution at 540 nm ± 10 nm in a microtiter plate reader (spectrophotometer), using the blanks as a reference.

Note: A mean $OD_{540 \pm 10nm}$ of 0.043 - 0.059 for the VC blanks is a target range of ODs but not a test acceptance criterion (range = mean OD ± 2.5 standard deviations; mean = 0.054; SD = 0.003; N = 114). Save raw data in the Microsoft Excel® template.

Note: The range of linearity of the microplate reader should be confirmed, as per in-house standard operating procedures. Additionally, all equipment should be calibrated according to manufacturer's instructions.

6.5.5 Quality Check of NHK NRU Assay

6.5.5.1 *Quality Check for PC*

- All acceptance criteria must be met by the PC for a test to be acceptable.
  - The PC (SLS) IC_{50} must be within ± two and a half (2.5) standard deviations (SD) of the historical mean established by the Test Facility and must have an r^2 (coefficient of determination) value calculated for the Hill model fit (i.e., from PRISM® software) ≥0.85. NICEATM/ECVAM study generated the following PC data:

    IC_{50} mean = 3.11 µg/mL; SD = 0.72 (n = 114)

    Range for IC_{50} mean ± 2.5 SD = 1.31 µg/mL – 4.91 µg/mL
  - The left and right mean of the VCs do not differ by more than 15% from the mean of all VCs.
  - At least one calculated cytotoxicity value >0% and ≤50% viability and at least one calculated cytotoxicity value >50% and <100% viability must be present.

6.5.5.2 *Quality Check for Test Substances*

- All acceptance criteria must be met by the test substances for a test to be acceptable.
  - The left and right mean of the VCs do not differ by more than 15% from the mean of all VCs.
  - At least one calculated cytotoxicity value >0% and ≤50% viability and at least one calculated cytotoxicity value >50% and <100% viability must be present.

Exception

If a test has only one point between 0 and 100% **and** the smallest dilution factor (i.e., 1.21) was used **and** all other test acceptance criteria were met, then the test will be considered acceptable.

Stopping Rule for Insoluble Substances

If the most rigorous solubility procedures have been performed and the assay cannot achieve adequate toxicity to meet the test acceptance criteria after three definitive trials, then the Study Director may end all testing for that particular substance.

Note: A corrected mean $OD_{540 \pm 10nm}$ of 0.205 – 1.645 for the VCs is a target range of ODs but not a test acceptance criterion (range = mean OD ± 2.5 standard deviations; mean = 0.685; SD = 0.175; N = 114).

6.5.3.3 *Checks for Systematic Cell Seeding Errors*

- To check for systematic cell seeding errors, untreated VCs are placed both at the left side (row 2) and the right side (row 11 for the test plates) of the 96-well plate. Aberrations in the cell monolayer for the VCs may reflect a volatile and toxic test substance present in the assay. If volatility is suspected, then proceed to **Section 6.5.6**.
- Checks for cell seeding errors may also be performed by examining each plate under a phase contrast microscope to assure that cell quantity is consistent.

6.5.6 Testing Volatile Substances

Although this test method is not suitable for highly volatile substances, mildly volatile substances may be tested with some success. Volatile test substances may generate vapors from the treatment medium during the test substance treatment incubation period. These vapors may become resorbed into the treatment medium in adjacent wells, such that culture wells nearest the highest doses may become contaminated by exposure. If the test substance is particularly toxic at the doses tested, the cross contamination may be

evident as a significant reduction in viability in the VC cultures (i.e., VC1) adjacent to the highest test substance doses.

If potential test substance volatility is suspected (e.g., for low density liquids) or if the initial range finder test (non-sealed plate) results show evidence of toxic effects in the control cultures (i.e., >15% difference in viability between VC1 [column 2] and VC2 [column 11]), then seal the subsequent test plates using the following procedure.

- Plates and substances will be prepared as usual according to **Sections 6.4** and **6.5**.
- Immediately after the 96-well culture plate has been treated with the suspected volatile substance (**Section 6.5.2**), apply the adhesive plate sealer (e.g., using a hand, microplate roller, etc.) directly over the culture wells. Assure that the sealer adheres to each culture well (well tops should be dry).
- Place the 96-well plate cover over the sealed plate and incubate the plate under specified conditions (**Section 6.5.2**). Note: Do not jam the plate lid over the film to avoid deforming the sealer and causing the sealer to detach from culture wells. Loose fit of the plate lid is acceptable.
- At the end of the treatment period, the plate sealer should be carefully removed to avoid spillage. Continue with the NRU assay as per **Section 6.5.4**.

6.6 Data Analysis

- The Study Director will use good biological/scientific judgment for determining *unusable* wells that will be excluded from the data analysis and provide explanations for the removal of any data from the analysis.
- A calculation of cell viability expressed as NRU is made for each concentration of the test substance by using the mean NRU of the six replicate values (minimum of four acceptable replicate well) per test

concentration (blanks will be subtracted). This value is compared with the mean NRU of all VC values. Relative cell viability is then expressed as percent of untreated VC. If achievable, the eight concentrations of each substance tested will span the range of no effect up to total inhibition of cell viability.

- Data from the microtiter plate reader should be transferred to a spreadsheet template (e.g., Microsoft Excel®) that will automatically determine cell viability, calculate IC_{50} values by linear interpolation, and perform statistical analyses (including statistical identification of outliers) (see **ANNEX III** for an example spreadsheet template).
- A Hill function analysis should be performed using statistical software (e.g., GraphPad PRISM® 3.0) and a template to calculate IC_{20}, IC_{50}, and IC_{80} values (and the associated confidence limits) for each test substance. The Hill function is recommended because all the dose-response information rather than a few points around the IC_{50} can be used to calculate the data. Additionally, the slope of the curve can be assessed using the Hill function.
- Dose-responses for which the toxicity plateaus as concentration increases do not fit the Hill function well when Bottom = 0. To obtain a better model fit, unconstrain the Bottom parameter so that the model calculates the Bottom value. However, when Bottom ≠ 0, the EC_{50} reported by the Hill function ≠ 50% viability since the Hill function defines EC_{50} as the point midway between Top and Bottom. To obtain the appropriate IC_{50} when Bottom ≠ 0, use the following rearranged Hill function:

$$X = \log EC_{50} - \frac{\log\left(\frac{Top - Bottom}{Y - Bottom} - 1\right)}{HillSlope}$$

X = the logarithm of concentration at 50% response, $\log EC_{50}$ = logarithm of concentration at the response midway between Top and Bottom,

Top = the maximum response, Bottom = the minimum response, Y = 50 (i.e., 50% response), and HillSlope = the steepness of the curve.

Note: IC_{50} values are used in a regression formula to predict the LD_{50} value of a test substance as an estimate of the starting dose for an acute oral toxicity test.

7.0 BACKGROUND REFERENCE MATERIALS

Hackenberg, U. and H. Bartling. 1959. Messen und Rechnen im pharmakologischen Laboratorium mit einem speziellen Zahlensystem (WL24-System). Arch. Exp. Pathol. Pharmakol. 235: 437-463.

Spielmann, H., S. Gerner, S. Kalweit, R. Moog, T. Wirnserberger, K. Krauser, R. Kreiling, H. Kreuzer, N.P. Luepke, H.G. Miltenburger, N. Müller, P. Murmann, W. Pape, B. Siegmund, J. Spengler, W. Steiling, and F.J. Wiebel. 1991. Interlaboratory assessment of alternatives to the Draize eye irritation test in Germany. Toxicol. *In Vitro* 5: 539-542.

Test Method Protocol for Solubility Determination. In Vitro Cytotoxicity Validation Study. Phase III. August 29, 2003. Prepared by The National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). Available: http://iccvam.niehs.nih.gov/methods/invitro.htm [accessed 15 September 2006].

U. S. Environmental Protection Agency. 1996. Product Properties Test Guidelines. OPPTS 803.7840. Water Solubility: Column Elution Method; Shake Flask Method. EPA712-C-96-041, Prevention, Pesticides and Toxic Substances, Washington DC.

Test Method Protocol for the 3T3 Neutral Red Uptake Cytotoxicity Test. A Test for Basal Cytotoxicity for an In Vitro Validation Study. November 4, 2004. Prepared by the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative

Toxicological Methods (NICEATM). Available:
http://iccvam.niehs.nih.gov/methods/invitro.htm [accessed 15 September 2006].

8.0. REFERENCES

Hartung T, Balls M, Bardouille C, Blanck O, Coecke S, Gstrauthaler G, and Lewis D. 2002. Good Cell Culture Practice: ECVAM Good Cell Culture Practice Task Force Report 1. Altern Lab Anim 30:407-414. Available:
http://ecvam.jrc.it/publication/index5007.html [accessed 02 September 2005].

ANNEX I

TEST METHOD PROCEDURE

Prequalification of Normal Human Epidermal Keratinocyte Growth Medium

This annex provides the guidelines and testing requirements for prequalifying manufacturer lots of Keratinocyte Basal Medium and the medium supplements for use with the Test Method Protocol for the NHK Neutral Red Uptake (NRU) Cytotoxicity Test. The medium and supplements should be tested so as to demonstrate their ability to perform adequately in the recommended assay.

The Testing Facility should request the quality control (QC) test data from the manufacturer for each potential lot of medium and supplements. Based upon the QC test data, purchase and test the one or two most current lots of medium and supplements that appear to have the potential to support NHK cultures according to the requirements of the aforementioned protocol.

1.0 TEST SYSTEM

The NHK NRU test is performed to analyze NHK growth characteristics and the *in vitro* toxicity of sodium lauryl sulfate (SLS), as measured by the IC_{50}, with each NHK medium/supplements being tested.

Every combination of medium/supplements expected to be used should be tested. Potential medium testing/supplement combinations are:

- One lot of medium/one lot of supplements: Test the lot of medium using the lot of supplements.
- Two or more lots of medium/one lot of supplements: Test each lot of medium using the one lot of supplements.
- One lot of medium/two or more lots of supplements: Test the lot of medium using each lot of supplements.

NHK cultures should be established using each medium/supplement combination to be tested, and should be subcultured on three different days into 96-well plates (1 plate per day) for three subsequent SLS cytotoxicity tests using each test medium/supplement combination along with a control medium (if available) for which performance has been previously established.

2.0 PROCEDURES

Prequalification of the keratinocyte medium and supplements will follow all procedures in the NHK NRU protocol.

2.1 Materials

See Section 6.1 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method.

2.2 Preparation of Media and Solutions

See Section 6.2 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method Protocol.

2.3 Methods

See Section 6.3 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method.

- NHK cultures should be established with cryopreserved cells seeded into individual tissue culture 25 cm^2 flasks using a proven medium/supplement combination (i.e., the control medium) and each test medium/supplement combination.
- Suspend freshly thawed cells initially into 9 mL of control medium and then add the cell suspension to 25 cm^2 culture flasks containing pre-warmed control or test medium. Cell seeding densities

(1 flask/density/medium) of 1 x 10^4, 5 x 10^3, and 2.5 x 10^3 are recommended.

- The cells should be subcultured on three different days into 96-well plates (see **Table C2-3**) for three subsequent NRU tests (three test plates total [one plate per day] for each medium/supplement combination and each control).

Table C2-3 Subculture Protocol

Flask	Subculture: 1 Test Plate and 1 Control Plate	Application of SLS
#1 (1 x 10^4 cells/mL)	Day A	Day X
#2 (5 x 10^3 cells/mL)	Day B	Day Y
#3 (2.5 x 10^3 cells/mL)	Day C	Day Z

- Subculturing the cells and application of the SLS will follow procedures in the protocol in reference to appropriate cell confluency. Cell numbers should be recorded for each flask prior to subculturing to the 96-well plates.

Note: Use of a control medium assumes that the Testing Facility has recent experience with a medium/supplement combination proven to support adequate NHK growth and provide adequate sensitivity to SLS. It is not absolutely necessary to use a control medium.

2.4 Doubling Time

See Section 6.3.5 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method. A doubling time experiment may be considered as an additional quality assurance check.

2.5 Preparation of SLS

See Section 6.4.1 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method.

- Preparation of SLS concentrations/dilutions should follow the main experiment (definitive assay) procedures specifically for testing compounds in Routine Culture Medium as outlined in Section 6.4.3 of the Test Method Protocol for the NHK NRU Cytotoxicity Test Method.
- The concentrations/dilutions should be the same or similar to those used previously with control medium/supplements.
- SLS concentration ranges used by three laboratories in the NICEATM/ECVAM validation study were 20.0 μg/mL – 1.4 μg/mL and 10.0 μg/mL – 0.6 μg/mL.

2.6 Test Procedure

See Sections 6.5.1, 6.5.2, and 6.5.4 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method.

- The C_1 test concentration will be the highest SLS concentration and C_8 the lowest concentration.
- Cells cultured in control medium and in each test medium/supplement combination should be tested in parallel for their sensitivity to SLS (see **Annex I, Section 2.3**).
- Each of the three test plates of the new medium/supplement combinations is considered a replicate test plate.

2.7 Microscopic Evaluation

See Section 6.5.3 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method.

Changes in morphology of the cells due to cytotoxic effects of the SLS (prior to measurement of NRU) should be recorded as per procedures outlined in Section 6.5.3 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method. In addition to the general microscopic evaluation of the cell cultures, the Study Director should make the following specific observations:

- *General culture observations*
  - rate of proliferation (e.g., rapid, fair, slow)
  - percent confluence (e.g., daily estimate)
  - number of mitotic figures (e.g., average per field)
  - contamination (present/not present)

- *Cell morphology observations*
  - overall appearance (e.g., good, fair, poor)
  - colony formation (e.g., tight/defined, fair, loose/migrating)
  - distribution (e.g., even/uneven)
  - abnormal cells (e.g., enlarged, vacuolated, necrotic, spotted, blebby - [average per field])

2.8 Data Analysis and Test Evaluation

See Section 6.6 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method. Test Acceptance Criteria in Section 6.5.5 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method should be used to determine acceptability of a test plate. Other criteria that should be considered by the Study Director includes the following:

- Mean corrected $OD_{540\text{-}550}$ of the VCs. Note: The target range for corrected mean $OD_{540 \pm 10nm}$ = 0.248 - 1.123 for the VCs, but it is not a test acceptance criterion (range = mean OD ± 2.5 standard deviations; mean = 0.685; SD = 0.175; N = 114).

- Cell morphology and confluence of the VCs at the end of the 48 hour treatment
- Doubling time

The Study Director should utilize all observed growth characteristics and test results in addition to comparison of results to the media manufacturer's QC data to determine whether the medium/supplements combinations perform adequately. The Testing Facility should request that the manufacturer reserve a portion of an acceptable lot based on estimates of media need.

ANNEX II

TEST METHOD PROCEDURE

Solubility Determination of Test Substances

1.0 PROPOSAL

This procedure was designed to identify the solvent that would provide the highest soluble concentration of a test substance so there would be uniform availability of the substance to cells used for *in vitro* basal cytotoxicity testing. The solubility exercises can be performed in a routine and repeatable manner and provide guidelines to effectively prepare test substances for toxicity testing in the Neutral Red Uptake (NRU) test methods. All individuals involved in solubility assessments should be trained so as to understand solvent and solubility issues.

2.0 TEST SYSTEM

The solubility test procedure is based on attempting to dissolve substances in various solvents with increasingly rigorous mechanical techniques. The solvents to be used, in the order of preference, are cell culture medium, dimethyl sulfoxide (DMSO), and ethanol (ETOH). Determination of whether a test substance has dissolved can be based on visual observation or through the use of a microscope. A test substance has dissolved if the solution is clear and shows no signs of cloudiness or precipitation.

3.0 PROCEDURES

3.1 Materials

See Section 6.1 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method.

3.2 Preparation of Media and Solutions

See Section 6.2 of Test Method Protocol for the NHK NRU Cytotoxicity Test Method. All solutions glassware, pipettes, etc., should be sterile and all procedures should be carried out under aseptic conditions and in the sterile environment of a laminar flow cabinet (biological hazard standard). All methods and procedures should be adequately documented.

3.3 Determination of Solubility

- Solubility should be determined in a step-wise procedure that involves attempting to dissolve a test substance at a relatively high concentration with the sequence of mechanical procedures specified in **Annex II, Section 3.5**. **Table C2-4** and **Figures C2-2** and **C2-3** illustrate the step-wise procedures.
- The hierarchy of preference of solvent for dissolving test substances is medium, DMSO, and then ETOH. If the substance does not dissolve in the solvent, the volume of solvent is increased so as to decrease the test substance concentration by a factor of 10, and then the sequence of mechanical procedures are repeated in an attempt to solubilize the substance at the lower concentrations.
- For testing solubility in medium, the starting concentration is 200,000 μg/mL (i.e., 200 mg/mL) in Tier 1, but for DMSO and ETOH the starting concentration is 200,000 μg/mL (i.e., 200 mg/mL) in Tier 3.

Table C2-4 Determination of Solubility in Routine Culture Medium, DMSO, or ETOH

Tier	1	2	3	4	5	6
Total Volume Routine Culture Medium	0.5 mL	0.5 mL	5 mL	50 mL		
Concentration of Test Substance Tier 1: Add ~ 100 mg to a tube. Add enough medium to equal Tier 1 volume. If insoluble, go to Tier 2. Tier 2: Add ~10 mg to another tube. Add enough medium to equal the first volume. Dilute to subsequent volumes if necessary.	200,000 µg/mL (200 mg/mL)	20,000 µg/mL (20 mg/mL)	2,000 µg/mL (2 mg/mL)	200 µg/mL (0.20 mg/mL)		
Total Volume DMSO/ETOH			0.5 mL	5 mL	50 mL	
Concentration of Test Substance (Add ~100 mg to a large tube. Add enough DMSO or ETOH to equal the first volume. Dilute with subsequent volumes if necessary.)			200,000 µg/mL (200 mg/mL)	20,000 µg/mL (20 mg/mL)	2,000 µg/mL (2 mg/mL)	
Total Volume DMSO/ETOH						50 mL
Concentration of Test Substance (Add ~10 mg to a large tube. Add enough DMSO or ETOH to equal 50 mL.)						200 µg/mL (0.2 mg/mL)
EQUIVALENT CONCENTRATION ON CELLS	100,000 µg/mL (100 mg/mL)	10,000 µg/mL (10 mg/mL)	1000 µg/mL (1 mg/mL)	100 µg/mL (0.1 mg/mL)	10 µg/mL (0.01 mg/mL)	1 µg/mL (0.001 mg/mL)

Abbreviations: DMSO: Dimethyl sulfoxide; ETOH: Ethanol.
Note: The amounts of test substance weighed and Routine Culture Medium added may be modified from the amounts given above, provided that the targeted concentrations specified for each tier are tested.

Figure C2-2 Solubility Step-Wise (Tiered) Procedure

TIER 1

STEP 1:	200 mg/mL test substance (TS) in 0.5 mL Routine Culture Medium • if TS soluble in medium, then **STOP**. • if TS insoluble in medium, then go to STEP 2.

TIER 2

STEP 2:	20 mg/mL TS in 0.5 mL Routine Culture Medium • if TS soluble, then **STOP**. • if TS insoluble, then go to STEP 3.

TIER 3

STEP 3:	200 mg/mL TS in DMSO • if TS soluble, then **STOP**. • if TS insoluble, test at 200 mg/mL in ETOH. – if TS soluble, then **STOP.** – If TS insoluble, go to STEP 4.

TIER 4

STEP 4:	0.2 mg/mL TS in medium (one or both) – increase volume from STEP 2 by 10 (i.e., to 50 mL) • if TS soluble in both media, then **STOP**. • if TS insoluble in one medium, test at 20 mg/mL in DMSO – increase volume from STEP 3 by 10 (i.e., to 5 mL). – if TS soluble, then **STOP.** – if TS insoluble, test at 20 mg/mL in ETOH – increase volume from STEP 3 by 10 (i.e., to 5 mL). ▪ if TS soluble, then **STOP.** ▪ if TS insoluble, then go to STEP 5.

TIER 5

STEP 5:	2 mg/mL TS in DMSO – increase volume from STEP 4 by 10 (i.e., to 50 mL) • if TS soluble, then **STOP.** • if TS insoluble, test at 2 mg/mL in ETOH – increase volume from STEP 4 by 10 (i.e., to 50 mL). – if TS soluble, then **STOP.** – if TS insoluble, then go to STEP 6.

TIER 6

STEP 6:	0.2 mg/mL TS in 50 mL DMSO • if TS soluble, then **STOP.** • if TS insoluble, test at 0.2 mg/mL in 50 mL ETOH – **STOP**

Abbreviations: DMSO: Dimethyl sulfoxide; ETOH: Ethanol.

Figure C2-3 Solubility Flow Chart

Tier	1		2		3		4		5		6
Concentration in Medium	**Start Here** 200 mg/mL	Incomplete solubility →	20 mg/mL	Incomplete solubility →	2 mg/mL		0.20 mg/mL				
					↓ Incomplete solubility		↓ Incomplete solubility				
Concentration in DMSO					200 mg/mL		20 mg/mL		2 mg/mL		0.2 mg/mL
					↓ Incomplete solubility		↓ Incomplete solubility		↓ Incomplete solubility		↓ Incomplete solubility
Concentration in ETOH					200 mg/mL	→ Incomplete solubility	20 mg/mL	→ Incomplete solubility	2 mg/mL	→ Incomplete solubility	0.2 mg/mL **End**
Concentration on Cells	100 mg/mL		10 mg/mL		1 mg/mL		0.1 mg/mL		0.01 mg/mL		0.001 mg/mL

Notes: NHK medium - Keratinocyte Growth Medium (e.g., KGM® from Cambrex) for normal human keratinocytes.

3.4 Methods

3.4.1 Tier 1

- Tier 1 begins with testing 200 mg/mL in Routine Culture Medium (see **Table C2-4**).
  - Weigh approximately 100 mg (100,000 μg) of the test substance into a glass tube. Document the test substance weight.
  - Add approximately 0.5 mL of medium into the tube so that the concentration is 200,000 μg/mL (200 mg/mL).
  - Mix the solution as specified in **Annex II, Section 3.5**. If complete solubility is achieved, then additional solubility procedures are not needed.

3.4.2 Tier 2

- If the test substance is insoluble in Tier 1 at 200 mg/mL, then proceed to Tier 2.
  - Weigh approximately 10 mg (10,000 μg) of the test substance into a glass tube. Document the substance weight.
  - Add approximately 0.5 mL of medium into the tube so that the concentration is 20,000 μg/mL (20 mg/mL).
  - Mix the solution as specified in **Annex II, Section 3.5**. If complete solubility is achieved, then additional solubility procedures are not needed.

3.4.3 Tier 3

- If the test substance is insoluble in Routine Culture Medium, proceed to Tier 3.
  - Add enough medium, approximately 4.5 mL, to attempt to dissolve the substance at 2 mg/mL by using the sequence of mixing procedures. If the test substance dissolves in medium at 2 mg/mL, no further procedures are necessary.
  - If the test substance does not dissolve in medium, weigh out approximately 100 mg test substance in a second glass tube and add

enough DMSO to make the total volume approximately 0.5 mL (for 200 mg/mL) and mix the solution as specified in **Annex II, Section 3.5**.

- If the test substance does not dissolve in DMSO, weigh out approximately 100 mg test substance in another glass tube and add enough ETOH to make the total volume approximately 0.5 mL (for 200 mg/mL) and mix the solution as specified in **Annex II, Section 3.5**.
- If the substance is soluble in either solvent, no additional solubility procedures are needed.

3.4.4 Tier 4

- If the substance is not soluble in Routine Culture Medium, DMSO, or ETOH at Tier 3, then continue to Tier 4 in **Table C2-4**.
  - Add enough solvent to increase the volume of the three (or four) Tier 2 solutions by 10 and attempt to solubilize again using the sequence of mixing procedures. If the test substance dissolves, no additional solubility procedures are necessary.
  - If the test substance does NOT dissolve, continue with Tier 5 and, if necessary, Tier 6 using DMSO and ETOH.

3.4.5 Tier 5

- Tier 5 begins by diluting the Tier 4 samples with DMSO or ETOH to bring the total volume to 50 mL. The mixing procedures are again followed to attempt to solubilize the substance.

3.4.6 Tier 6

- Tier 6 is performed, if necessary, by weighing out another two samples of test substance at ~10 mg each and adding ~50 mL DMSO or ETOH for a 200 μg/mL solution, and following the mixing procedures.

Example

- If complete solubility is not achieved at 20,000 μg/mL in Routine Culture Medium at Tier 2 using the mixing procedures, then the procedure continues to Tier 3 by diluting the solution to 5 mL with medium and mixing again.
- If the substance is not soluble in Routine Culture Medium, two samples of ~ 100 mg test substance are weighed to attempt to solubilize in DMSO and ETOH at 200,000 μg/mL (i.e., 200 mg/mL). Solutions are mixed following the sequence of procedures prescribed in **Annex II, Section 3.5** in an attempt to dissolve.
- If solubility is not achieved at Tier 3, then the solutions prepared in Tier 3 are diluted by 10 so as to test 200 μg/mL in media, and 20,000 μg/mL in DMSO and ETOH. This advances the procedure to Tier 4. Solutions are again mixed in an attempt to dissolve.
- If solubility is not achieved in Tier 4, the procedure continues to Tier 5, and to Tier 6 if necessary (see **Figures C2-2** and **C2-3** and **Table C2-4**).

3.5 Mechanical Procedures

The following hierarchy of mixing procedures will be followed to dissolve the test substance:

- Add test substance to solvent as in Tier 1 of **Table C2-4**. (Test substance and solvent should be at room temperature.)
- Gently mix at room temperature. Vortex the tube (1 –2 minutes).
- If test substance has not dissolved, use waterbath sonication for up to 5 minutes.
- If test substance is not dissolved after sonication, then warm solution to 37 °C for 5 - 60 minutes. This can be performed by warming tubes in a 37 °C waterbath or in a CO_2 incubator at 37 °C. The solution may be stirred during warming (stirring in a CO_2 incubator will help maintain proper pH).
- Proceed to Tier 2 (and Tiers 3-6, if necessary of **Table C2-4** and repeat procedures 2-4).

The preference of solvent for dissolving test substances is Routine Culture Medium, DMSO, and then ETOH. Thus, if all solvents for a particular tier are tested simultaneously and a test substance dissolves in more than one solvent, then the choice of solvent follows this hierarchy. For example, if, at any tier, a substance were soluble in Routine Culture Medium and DMSO, the choice of solvent would be medium. If the substance were insoluble in medium, but soluble in DMSO and ETOH, the choice of solvent would be DMSO.

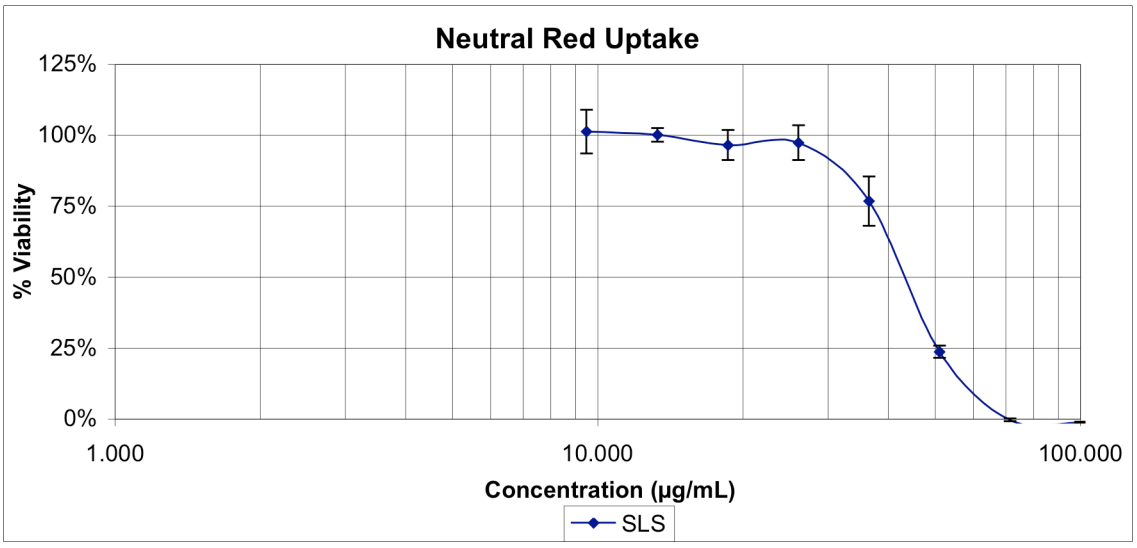
[This Page Intentionally Left Blank]

ANNEX III

Microsoft EXCEL® Example Spreadsheet Template

Test Facility :	A					Study Number.:	A1					
Chemical Code :	SLS					96-Well Plate ID :	A11					
2nd Chem. Code*:	11					Experiment ID :	XX					
					96-WELL PLATE MAP							
	1	2	3	4	5	6	7	8	9	10	11	12
A	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank
B	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
C	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
D	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
E	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
F	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
G	Blank	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	Blank
H	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank	Blank
					RAW ABSORBANCE DATA (OD_{550})							
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.044	0.044	0.045	0.045	0.045	0.046	0.051	0.057	0.057	0.043	0.041	0.044
B	0.042	0.456	0.043	0.043	0.130	0.300	0.395	0.414	0.418	0.402	0.401	0.042
C	0.043	0.407	0.042	0.041	0.130	0.294	0.383	0.382	0.413	0.375	0.385	0.044
D	0.043	0.438	0.042	0.043	0.147	0.337	0.409	0.404	0.438	0.436	0.391	0.047
E	0.044	0.448	0.041	0.045	0.132	0.321	0.429	0.414	0.416	0.420	0.441	0.042
F	0.045	0.411	0.040	0.042	0.127	0.375	0.397	0.402	0.422	0.447	0.403	0.043
G	0.041	0.405	0.043	0.040	0.124	0.361	0.444	0.442	0.425	0.448	0.405	0.044
H	0.041	0.041	0.048	0.042	0.042	0.044	0.042	0.042	0.040	0.044	0.041	0.041
Max	0.045	0.456	0.043	0.045	0.147	0.375	0.444	0.442	0.438	0.448	0.441	0.047
Min	0.041	0.405	0.040	0.040	0.124	0.294	0.383	0.382	0.413	0.375	0.385	0.041
Next Max	0.044	0.448	0.042	0.043	0.132	0.361	0.429	0.414	0.425	0.447	0.405	0.044
Next Min	0.042	0.407	0.041	0.041	0.127	0.300	0.395	0.402	0.416	0.402	0.391	0.042
Rmax	-0.250	-0.157	-0.333	-0.400	-0.652	-0.173	-0.246	-0.467	-0.520	-0.014	-0.643	-0.500
Rmin	0.250	0.039	0.333	0.200	0.130	0.074	0.197	0.333	0.120	0.370	0.107	0.167
					CORRECTED ABSORBANCE (Sample OD_{550} - Mean Blank OD_{550})							
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.001	0.001	-0.002	0.002	0.002	0.001	0.005	0.008	0.009	-0.001	-0.002	0.001
B	-0.001	0.413	-0.004	-0.001	0.087	0.255	0.349	0.365	0.370	0.359	0.358	-0.001
C	0.000	0.364	-0.005	-0.003	0.087	0.249	0.337	0.333	0.365	0.332	0.342	0.001
D	0.000	0.395	-0.005	-0.001	0.104	0.292	0.363	0.355	0.390	0.393	0.348	0.004
E	0.001	0.405	-0.006	0.002	0.089	0.276	0.383	0.365	0.368	0.377	0.398	-0.001
F	0.002	0.368	-0.007	-0.001	0.084	0.330	0.351	0.353	0.374	0.404	0.360	0.000
G	-0.002	0.362	-0.004	-0.004	0.081	0.316	0.398	0.393	0.377	0.405	0.362	0.001
H	-0.002	-0.002	0.002	-0.001	-0.001	-0.001	-0.005	-0.008	-0.009	0.001	-0.002	-0.002
Mean Blank =	0.043		0.047	0.044	0.044	0.045	0.047	0.050	0.049	0.044		
					RELATIVE VIABILITY (% OF VEHICLE CONTROL)							
	1	2	3	4	5	6	7	8	9	10	11	12
A												
B		110.7%	-0.9%	-0.1%	23.2%	68.4%	93.4%	97.7%	99.0%	96.1%	96.0%	
C		97.6%	-1.2%	-0.7%	23.2%	66.7%	90.2%	89.1%	97.7%	88.9%	91.7%	
D		105.9%	-1.2%	-0.1%	27.7%	78.3%	97.2%	95.0%	104.4%	105.2%	93.3%	
E		108.6%	-1.5%	0.4%	23.7%	74.0%	102.5%	97.7%	98.5%	100.9%	106.7%	
F		98.7%	-1.7%	-0.4%	22.4%	88.5%	94.0%	94.5%	100.1%	108.2%	96.5%	
G		97.1%	-0.9%	-0.9%	21.6%	84.7%	106.5%	105.2%	100.9%	108.4%	97.1%	
H												

Test Facility :	A					Study Number.:	A1				
Chemical Code :	SLS					96-Well Plate ID :	A11				
2nd Chem. Code*:	11					Experiment ID :	XX				
	VC1	C1	C2	C3	C4	C5	C6	C7	C8	VC2	
Conc. (µg/mL) :	0.0	100	71.4	51.0	36.4	26.0	18.6	13.3	9.49	0.0	
Mean Corr. OD :	0.385	-0.005	-0.001	0.088	0.286	0.363	0.360	0.374	0.378	0.361	
SD :	0.023	0.001	0.002	0.008	0.033	0.023	0.020	0.009	0.029	0.020	
Mean Vehicle Control :	0.373										
Mean Blank :	0.043										
% of Vehicle Control :	103.1%	-1.3%	-0.3%	23.6%	76.8%	97.3%	96.5%	100.1%	101.3%	96.9%	
SD :	6.0%	0.3%	0.5%	2.1%	8.7%	6.2%	5.3%	2.4%	7.7%	5.2%	
% CV :	5.86%	-25.1%	-150%	9.09%	11.4%	6.33%	5.47%	2.39%	7.59%	5.41%	
Mean VC - VC1 (%) :	-3.1%										
Mean VC - VC2 (%) :	3.1%										
Mean Absolute OD :	0.416										
					Visual Observations						
	VC	C1	C2	C3	C4	C5	C6	C7	C8		
ENTER CODES:	1	4	4	3	2	1	1	1	1		
				Interpolated IC_{50} :		**4.32E+01**	**µg/mL**				



TEST CHEMICAL

Test Facility :	A	Study Number.:	A1
Chemical Code :	SLS	96-Well Plate ID :	A11
2nd Chem. Code*:	11	Experiment ID :	XX

* Testing Facility Accession Code, if applicable

PREPARATION OF TEST CHEMICAL

Solvent:	Medium	Dilution factor:	1.4	
Solvent Conc. (%, v/v) in dosing solutions :	N/A	Highest Stock Conc.:	20,000	µg/mL

Aids used to dissolve : ☐ Vortexing ☐ sonication ☐ heating to 37C

pH (highest medium stock or 2X dosing solution) : 8.0

Medium Clarity/Color (highest 2X dosing solution): clear red If ppt, note lowest conc.:

Concentration Series (µg/mL)

C1	C2	C3	C4	C5	C6	C7	C8
100	71.4	51.0	36.4	26.0	18.6	13.3	9.49

Positive Control (SLS) 100 - 9.49 µg/mL

CELL LINE/TYPE

Name:	BALB/c 3T3	Supplier:	ATCC	Lot No.	not provided
Passage No.:	69	Passage No. in Assay:	75	Proliferating/frozen	24-May-02

CELL CULTURE CONDITIONS

Medium:	DMEM	Supplier:		Lot No.:	
Serum:	NCS	Supplier:		Lot No.:	
Serum Conc.:		Growth Medium:	10%	Treatment Medium:	0%

TEST ACCEPTANCE CRITERIA

No. of values >50% and <100%:	3	No. of values >0% and ≤50%:	1	Accept?	YES
VC: % Difference between Col 2 and mean VC.:			-3%	Accept?	YES
PC: Hill Function R^2 Value of SLS:		0.99		Accept?	YES
PC: IC_{50} of SLS:		43.2	µg/mL	Accept?	YES

TIMELINE

Cell Seeding Date	Dose Application Date	OD_{550} Determination Date

TEST RESULTS

VC: Mean Corrected OD_{550}: 0.373 Hill Function R^2 Value: 0.9869

log IC20 :	1.551E+00 µg/mL	log IC50 :	1.635E+00 µg/mL	log IC80 :	1.718E+00 µg/mL
IC20 :	3.56E+01 µg/mL	IC50 :	4.32E+01 µg/mL	IC80 :	5.22E+01 µg/mL

Test Chemical F.W. : 288.4

IC_{20} :	0.12331183 mM	IC_{50} :	0.1496252 mM	IC_{80} :	0.18113599 mM

[This Page Intentionally Left Blank]

APPENDIX D

FEDERAL REGISTER NOTICES AND PUBLIC COMMENTS

[This Page Intentionally Left Blank]

APPENDIX D1

FEDERAL REGISTER NOTICES

[This Page Intentionally Left Blank]

APPENDIX D2

ICCVAM CONSIDERATION OF PUBLIC COMMENTS RECEIVED IN RESPONSE TO FEDERAL REGISTER NOTICES

[This Page Intentionally Left Blank]

In response to eight *Federal Register* (*FR*) notices that were released between June 2000 and July 2006, 298 public comments were received. Comments received in response to the *FR* notices and/or were related to those *FR* notices can be obtained on CD ROM upon request to The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) by mail, fax, or email (NICEATM, NIEHS, P.O. Box 12233, MD EC-17, Research Triangle Park, NC 27709, (phone) 919-541-2384, (fax) 919-541-0947, (email) niceatm@niehs.nih.gov). The following sections, delineated by *FR* notice, provide a brief discussion of the public comments received in response to three of the published *FR* notices.

1.0 Public Comments Received in Response to *FR* Notice Released on March 22, 2005 (Volume 70, Number 54; pages 14473-14474)

NICEATM, in an *FR* notice (70 *FR* 54:14473-14474, March 22, 2005) requested nominations of scientific experts for consideration as part of an independent peer review panel to evaluate the validation status of two *in vitro* cytotoxicity assays for estimating *in vivo* oral toxicity. One comment was received in response to this request and stated that animal testing should be stopped and that more accurate test methods that are more humane should be used.

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) appreciates the comment received. It should be noted that ICCVAM does not determine whether a test method is acceptable for use by U.S. Federal agencies or the international regulatory community. ICCVAM develops and forwards recommendations on the usefulness and limitations of the proposed test methods to each U.S. Federal agency for its review. Based on their specific statutory mandates, each U.S. Federal agency will consider ICCVAM's recommendations and then make a determination as to the acceptability of the test methods.

2.0 Public Comments Received in Response to *FR* Notice Released on March 21, 2006 (Volume 71, Number 54; pages 14229-14231)

NICEATM, in an *FR* notice (71 *FR* 54:14229-14231, March 21, 2006) requested comments on (1) the draft Background Review Document (BRD) being forwarded to the Scientific Peer Review Panel, (2) the draft ICCVAM test method recommendations, (3) draft test method protocols, and (4) draft performance standards. In response to this *FR* notice, 297 comments were received.

Of the comments received, 296 comments stated that there was a consensus at the workshop in 2000 (*In Vitro* Methods for Assessing Acute Systemic Toxicity) that cell-based methods could be used immediately to reduce the number of animals killed and could potentially be validated as replacements to current acute systemic toxicity test methods, given the proper funding and effort. However, the comments stated that announcement for the Peer Review Panel meeting scheduled for 2006 did not mention the potential of using these cell-based methods as potential replacement methods.

ICCVAM considered all the recommendations from the workshop in developing its own recommendations for activities after the 2000 workshop. The ICCVAM recommendations were forwarded to U.S. Federal agencies, along with the workshop report and the *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity*. Consistent with the workshop recommendations, ICCVAM recommended that the near-term focus for validation should be on characterizing the usefulness of two standardized *in vitro* assays using rodent and human cells in predicting acute toxicity with a broader range of chemicals than had been previously tested. Therefore, the current evaluation focused on the use of these two *in vitro* methods for estimating starting doses for acute oral systemic toxicity tests.

Of the comments received, 23 stated that it was time to refine and implement non-animal, cell-based methods to replace current systemic acute toxicity test method protocols. ICCVAM appreciates the comments received. It should be noted that ICCVAM does not

determine whether a test method is acceptable for use by U.S. Federal agencies or the international regulatory community. ICCVAM develops and forwards recommendations on the usefulness and limitations of the proposed test methods to each U.S. Federal agency for its review. Based on their specific statutory mandates, each U.S. Federal agency considers ICCVAM's recommendations and then makes a determination as to the acceptability of the test methods.

Of the comments received, two focused on the rationale for ICCVAM to not consider or implement the recommendations of the participants of the *International Workshop on In Vitro Methods for Assessing Acute Systemic Toxicity*. ICCVAM notes that the participants of the workshop made the following recommendations (among others):

- *In vitro* cytotoxicity data should be used to predict starting doses for *in vivo* lethality studies.
- Test laboratories should evaluate and compare the performance of several *in vitro* cytotoxicity tests with the existing Registry of Cytotoxicity (RC) data.
- A prevalidation study should be initiated as soon as possible to evaluate various cell types, exposure periods, and endpoint measurements as predictors of acute toxicity. The assay, or battery of assays, determined to be the best predictor of *in vivo* lethality could then be optimized further to identify, standardize, and validate simple predictive systems for gut absorption, blood-brain barrier passage, kinetics, and metabolism.
- In the longer-term, preferably as a parallel activity, there should be a focus on the development and validation of human *in vitro* test systems for predicting human acute toxicity.
- The evaluation and ultimate acceptance of *in vitro* assays for human acute toxicity will need a larger reference database than is presently available for validation purposes.

ICCVAM considered these as well as other recommendations from the workshop in developing its own recommendations. The ICCVAM recommendations were forwarded to U.S. Federal agencies along with the workshop report and *Guidance Document on Using In*

Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity. Consistent with the workshop recommendations, ICCVAM recommended that the near-term focus for validation should be on characterizing the usefulness of two standardized *in vitro* assays using rodent and human cells in predicting acute toxicity with a broader range of chemicals than had been previously tested. The NICEATM/European Centre for the Validation of Alternative Methods (ECVAM) validation study was based on this recommendation and its goals and purpose are entirely consistent with the workshop recommendations. Research activities to identify appropriate *in vitro* absorption, distribution, metabolism, and excretion systems was identified as a longer-term objective. At the same time, NICEATM proceeded with a validation study to establish the utility of setting the starting dose across the range of Globally Harmonized System of Classification and Labelling of Chemicals (GHS) hazard classification; and to establish a high quality database as a foundation for the development of other *in vitro* tests that could be used, along with *in vitro* basal cytotoxicity test methods, to improve the prediction of *in vivo* acute toxicity.

ICCVAM received a comment that the objectives of the NICEATM/ECVAM validation study appeared to be a mixture of partly conflicting goals (e.g., validating the RC prediction model, assessing the boundaries of applicability, and assessing the predictive capacity of LD_{50} point measures). As stated in the BRD, ICCVAM notes that the validation study objectives were to:

- Further standardize and optimize two *in vitro* neutral red uptake (NRU) cytotoxicity protocols using mouse fibroblast (BALB/c) 3T3 cells and normal human epidermal keratinocytes (NHK) in order to maximize intra- and inter-laboratory reproducibility
- Refine the prediction model drawn from the German Center for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET) approach
- Assess the accuracy of the two standardized *in vitro* basal cytotoxicity test methods for estimating rodent LD_{50} values across the five GHS (UN 2005) categories of acute oral toxicity as well as unclassified toxicities and estimating human lethal serum concentrations

- Estimate the reduction and refinement in animal use achievable from using *in vitro* basal cytotoxicity assays as one of the factors of the weight-of-evidence to identify starting doses for specific rodent acute oral toxicity tests
- Generate high quality *in vivo* lethality and *in vitro* cytotoxicity databases that can be used to support the investigation of other *in vitro* test methods necessary to improve the prediction of acute systemic toxicity

ICCVAM received a comment focused on the selection of the test chemicals for the validation study. The comment noted that these chemicals were not appropriate to achieve the main goal of the validation study (i.e., verification or falsification of the RC prediction model). ICCVAM appreciates the comment but notes that the verification of falsification of the RC prediction model was not a goal of this effort (see above).

ICCVAM received a comment regarding the variability of *in vitro* data obtained during Phase I and Phase II of the validation study. The comment stated that the *in vitro* test protocols were optimized, and that the necessity of this step was questionable. The comment recommended that the outcomes from this study be compared with other interlaboratory validation studies that have used the 3T3 NRU standard protocol. ICCVAM notes that the test acceptance criteria for the vehicle control optical density and placement of the cytotoxicity points were revised after it was noted that good dose-response data were obtained even in tests that failed the original criteria. Thus, to increase the test method experimental success rate, the criteria were revised. These changes did not alter the performance of the test methods.

Regarding the variability of the *in vitro* data, this comment appears to refer to the difference between the 3T3 NRU and NHK NRU IC_{50} values since no such variation occurred across laboratories for the same cell type. ICCVAM notes that it should not be a surprise that, for some chemicals, large variation exists for IC_{50} results obtained using different cell lines even when using very similar test protocols. Such data are important for characterizing which cell line(s) may be optimal for *in vitro* cytotoxicity testing and for identifying chemicals that may require additional evaluation.

ICCVAM received a comment regarding the variability of the *in vivo* reference data. The comment noted that there had been extensive efforts by ICCVAM to obtain multiple *in vivo* LD_{50} data per test chemical. The comment noted that while most validation studies assess the variability of the *in vivo* data to analyze the performance of the alternative methods, this type of analysis was not present in the BRD. ICCVAM appreciates the comments and notes that the BRD analyzed the variation of *in vivo* data in Section 4. Table 4-2 in the BRD provides the ratio of the maximum to the minimum acceptable LD_{50} for each chemical.

ICCVAM received a comment stating that the evaluation of the two *in vitro* assays was highly biased by the unbalanced selection of chemicals used in the validation study. The commenter stated that all calculations (e.g., the contingency tables for prediction of the GHS classes) were influenced by the bias in the chemical selection, so that even the strength of the prediction model (correct prediction of the absence of toxicity) was lost. The commenter stated that a thorough discussion of the influence of chemical selection on the study outcome should be included.

ICCVAM agrees with the comment that the selection of chemicals and their fit to the regression being evaluated affects the accuracy of GHS category predictions. Even though the selection of chemicals and their fit to the regressions affects the accuracy of GHS category predictions, the analyses provide a valid comparison of the test methods to one another and of the regressions to one another.

A comment was received stating that the results of the current study should be correlated to the results and information obtained from previous studies. ICCVAM agrees and notes that Section 9 of the BRD provides a literature review of studies most relevant to the NICEATM/ECVAM validation study. The literature review addresses (a) the use of *in vitro* NRU cytotoxicity test methods for correlations with rodent lethality and other toxicities and (b) the use of *in vitro* basal cytotoxicity to predict starting doses for acute oral lethality assays.

ICCVAM received a comment related to (a) the draft ICCVAM recommendation proposing that the RC should be revised and (b) the draft minimum performance standards. ICCVAM appreciates the comment received and notes that the proposed revisions were based on a variety of factors, were independent of each other, and are justified based on the breadth of the RC database. Furthermore, ICCVAM notes that the draft performance standards take into account the technical aspects of the test methods and proposes reference substances compatible with the RC regression after excluding substances without rat LD_{50} data and those with known mechanisms of action.

3.0 Public Comments Received in Response to *FR* Notice Released on July 11, 2006 (Volume 71, Number 132; pages 39122-39123)

NICEATM, in an *FR* notice (71 *FR* 132:39122-39123, Jul 11, 2006) requested comments on the Panel's conclusions on the draft ICCVAM test method recommendations. In response to this *FR* notice, one comment was received.

The comment stated that there was concern that despite near unanimous agreement at the 2000 workshop that the cell-based methods could be used immediately to set the starting dose for oral toxicity tests and that given appropriate effort and funding these method could be validated as a replacement measure, there has been little progress on the issue. There was concern that the Peer Panel Report did not require the use of the *in vitro* methods to estimate a starting dose, due to the understandable contention that significant information may already be available on the chemical or its class. The commentor stated that companies should be encouraged to use the non-animal methods to obtain another level of comfort with using and reading data generated by them. The comment stated that, based on the available scientific evidence, the Peer Panel Report should address expedient steps to replace lethal dose animal tests at the extremes of toxicity.

ICCVAM appreciates the comments provided. ICCVAM notes that the Peer Panel Report contains the conclusions of the Peer Review Panel and the document would not be edited by ICCVAM. However, the Peer Panel Report and all the comments received in response to the

published *FR* notices were considered by ICCVAM during the development of the ICCVAM Test Method Evaluation Report.

APPENDIX E

ICCVAM RECOMMENDATIONS FROM THE 2000 INTERNATIONAL WORKSHOP ON *IN VITRO* METHODS FOR ASSESSING ACUTE SYSTEMIC TOXICITY

ICCVAM Recommendations on *In Vitro* Methods for Assessing Acute Systemic Toxicity[11]

An International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity was convened in Arlington, VA, on October 17-20, 2000. The Workshop was organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), and was co-sponsored by the U.S. Environmental Protection Agency (EPA) and the National Institute of Environmental Health Sciences (NIEHS), and the National Toxicology Program (NTP). The Workshop focused on reviewing the validation status and possible current uses of *in vitro* methods to assess acute oral lethality potential of chemicals. Workshop participants also recommended research, development, and validation efforts that would further advance the usefulness of *in vitro* methods. For a complete account of Workshop discussions and recommendations, please refer to the *Report of the International Workshop on In Vitro Methods for Assessing Acute Systemic Toxicity* (ICCVAM 2001a). Based on a review of the Workshop Report, ICCVAM developed the following recommendations that were forwarded to Federal agencies with the Report and Guidance Document.

Current Uses for *In Vitro* Methods

Workshop participants considered the merit of using *in vitro* cytotoxicity tests for predicting the acute oral lethality of chemicals in humans and animals, as suggested by previous studies (e.g., Clemedson and Ekwall, 1999; Halle and Goeres, 1988). They concluded that the available *in vitro* assays would require further development to accurately predict acute lethality (i.e., LD_{50}). Workshop participants recommended that *in vitro* cytotoxicity data be included as one of the factors used to identify appropriate starting doses for *in vivo* acute lethality studies as described by Spielmann et al. (1999). In the approach developed by Spielmann, *in vitro* cytotoxicity tests are used to predict starting doses for acute *in vivo* lethality assays.

ICCVAM agrees with the Workshop Report that data from *in vitro* cytotoxicity assays can be useful as one of the tools (e.g., SAR or bridging from similar compounds or mixtures) in setting a starting dose for the *in vivo* assessment of acute oral toxicity. The attached *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (ICCVAM 2001b) describes one method, the murine BALB/c 3T3 neutral red uptake assay, for which data for a number of chemicals supports its potential utility for estimating the starting dose. Starting doses are calculated using a regression formula based on an *in vitro-in vivo* correlation for 347 chemicals. Preliminary information suggests that use of this *in vitro* approach could reduce the number of animals currently used in *in vivo* acute toxicity tests. Additionally, new OECD Guidelines for *in vivo* acute toxicity testing

[11] ICCVAM. 2001a. Appendix I

recommend a starting dose below the estimated LD_{50} to minimize the number of animals that receive lethal doses and to avoid underestimating the hazard. ICCVAM recommends that Federal agencies consider making information about this *in vitro* approach available as one of the tools that can be used to select an appropriate starting dose for acute oral toxicity tests.

Research Directions

Workshop participants identified several areas for research and development activities to advance the use of *in vitro* methods for predicting acute oral toxicity in animals and humans. ICCVAM recognizes that there are many directions that such future research and testing might take. These include both near-term and long-term research activities.

A. Near-Term Research

ICCVAM concurs with the Workshop recommendation that near-term validation studies should focus on two standard cytotoxicity assays: one using a human cell system and one using a rodent cell system. Since the murine BALB/c 3T3 cytotoxicity assay has been evaluated for only a limited number of chemical classes, there is merit in determining its usefulness with a broader array of chemical classes. Cell lines established from the rat rather than the mouse might also be considered, as most acute oral toxicity testing is conducted in this species. Human cell lines should also be considered since one of the aims of toxicity testing is to make predictions of potential toxicity in humans. Future validation studies should therefore compare rodent and human *in vitro* data with one another, with rodent *in vivo* data, and with human *in vivo* data. Correlations between *in vitro* and *in vivo* data might help in selecting cytotoxicity assays for further evaluation.

The U.S. EPA and NIEHS are collaborating to further characterize the usefulness of *in vitro* methods for acute toxicity testing. ICCVAM recognizes that these activities may yield important information on the near-term and long-term application of *in vitro* tests. ICCVAM recommends the establishment of an interagency expert group under ICCVAM to advise on near-term activities such as assay selection, study design, and chemical selection.

- Long-Term Research

Longer-term research activities should be directed at improving *in vitro* systems that provide information on biokinetics, metabolism, and organ-specific toxicity. *In vitro* methodologies for gathering biokinetic and target organ specific effects data are needed to facilitate reasonably accurate predictions of LD50s, signs and symptoms associated with toxicity, and pathophysiological effects. Research efforts that might increase the predictive capability of *in vitro* assays include:

- Developing the use of quantitative structure-activity relationship (QSAR)/quantitative structure-property relationship (QSPR) models that predict kinetic parameters such as gut absorption and passage across the brain, kidney, and skin barrier systems.

- Developing efficient *in vitro* systems that provide accurate metabolic and biokinetic data.
- Developing accurate physiologically-based biokinetic models.
- Developing *in vitro* systems that accurately predict organ-specific toxicity.
- Investigating the mechanistic basis for "outlier" chemicals in *in vitro-in vivo* correlations and developing "exclusion" rules for identifying chemicals that cannot be accurately evaluated using *in vitro* methods.
- Investigating the utility of toxicogenomics/proteomics for the assessment of acute toxicity, especially the prediction of NOAELs/LOAELs for acute exposure.

ICCVAM appreciates that most of these long-term research activities will yield further improvements in the usefulness of *in vitro* methods for predicting acute systemic toxicity, but that significant resources would be required. ICCVAM concludes that such activities will warrant consideration along with other potential research efforts in establishing priorities.

REFERENCES

Clemedson, C., and B. Ekwall. 1999. Overview of the Final MEIC Results: I. The *In Vitro-In Vivo* Evaluation. Toxicol. *In Vitro* 13: 1-7.

Halle, W., and E. Goeres. 1988. Register der Zytotoxizität (IC50) in der Zellkultur und Möglichkeiten zur Abschätzung der akuten Toxizität (LD50). In: Beiträge zur Wirkstoffforschung, Institute für Wirkstoffforschung. (Oehme, P., H. Loewe, and E. Goeres, eds). Berlin, Germany.

Halle, W., and H. Spielmann. 1992. Two Procedures for the Prediction of Acute Toxicity (LD50) from Cytotoxicity Data. ATLA 20: 40-49.

ICCVAM. 2001a. Report Of The International Workshop On In Vitro Methods For Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].

ICCVAM. 2001b. Guidance Document On Using In Vitro Data To Estimate In Vivo Starting Doses For Acute Toxicity. NIH Publication No. 01-4500. National Institute for Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/. [accessed 2 June 2005].

Spielmann, H., E. Genschow, M. Leibsch, and W. Halle. 1999. Determination of the Starting Dose for Acute Oral Toxicity (LD50) Testing in the Up-and-Down Procedure (UDP) from Cytotoxicity Data. ATLA 27(6): 957-966.